ATTORNEY'S DOCKET NUMBER: 2007651-0001

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants:	Jack *et al.*	**Examiner:**	Hutson, Richard G.
Serial No.:	10/089,027	**Art Unit:**	1652
Filing Date:	March 26, 2002	**Confirmation No.:**	9409
Title:	INCORPORATION OF MODIFIED NUCLEOTIDES BY ARCHAEON ANA POLYMERASES AND RELATED METHODS		

Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

ELECTRONICALLY FILED VIA EFS WEB

Sir:

APPEAL BRIEF UNDER 37 C.F.R. § 41.37

Appellants hereby appeal to the Board of Patent Appeals and Interferences (the "Board") from the Examiner's final rejection of pending claims 32-42 of the above-referenced application.

A final Office Action was mailed on April 21, 2009. A Notice of Appeal was filed on July 21, 2009. This Appeal Brief is filed on December 23, 2009 with a Petition under 37 C.F.R. § 1.136 for four month extension of time. An electronic payment of the $270.00 fee for filing an appeal brief under 37 C.F.R 41.20(b)(2) and $865 fee for an extension of time is filed concurrently with this submission. Applicant believes that no further petitions and fees are required for this Appeal Brief to be entered. Please consider this a conditional petition for any additional extensions, if needed, and please charge any additional fees or credit any overpayments that may be required to our Deposit Account No. 03-1721 referencing attorney docket number 2007651-0001.

REAL PARTY IN INTEREST

As a result of an assignment by the inventors in the present application, the real party in interest in this application is New England Biolabs, Inc. The assignment to New England Biolabs, Inc. was recorded in the Patent and Trademark Office at Reel 012921, Frame 0103.

RELATED APPEALS AND INTERFERENCES

No other appeals or interferences are known to Appellants, Appellants' legal representative, or Appellants' assignee that will directly affect or be directly affected by the Board's decision in this appeal. Similarly, no such appeals or interferences are known that may have a bearing on the Board's decision in this appeal.

STATUS OF CLAIMS

The application was filed with 31 claims. Various claims were amended and/or cancelled in Amendments filed on June 7, 2002, April 18, 2005 (not entered), August 15, 2005, May 4, 2006 (not entered), May 30, 2006 (not entered), July 3, 2006, and February 24, 2007. Pending claims 2-4, 13-22, and 27-31 were canceled in the amendment filed October 31, 2007, and new claims 32-43 were presented. Claim 32 was amended in an Amendment filed on August 29, 2008 (not entered) and an Amendment filed on January 9, 2009. Claims 32-42 were finally rejected in an Office Action mailed April 21, 2009. Claim 43 is objected to for depending on rejected claims 32 and 33. The rejection of claims 32-42 is hereby appealed. A listing of pending claims 32-43 is provided in the attached **Claims Appendix**.

STATUS OF AMENDMENTS

There are no outstanding amendments to the claims.

SUMMARY OF CLAIMED SUBJECT MATTER

DNA polymerases are enzymes that catalyze polymerization of nucleotides into a DNA strand. The present invention encompasses the finding that a certain class of DNA polymerases has the ability to incorporate a particular modified type of nucleotide, acyclonucleotides, into DNA strands. The present claims therefore recite use of DNA polymerases from that class (defined by the present of a particular amino acid motif whose presence is shown to correlate with the activity) to incorporate acyclonucleotides into a polynucleotide chain.

Independent claim 32 and dependent claims 33-43 specifically recite methods comprising steps of providing a DNA polymerase of the relevant class, contacting the DNA polymerase with a template, a primer that binds to the template, and a collection of nucleotides including at least one acyclonucleotide, and incubating the DNA polymerase with the template and the nucleotides so that the DNA polymerase extends the primer by incorporating the nucleotides. The claims require that the utilized DNA polymerase be a member of the relevant class of DNA polymerases by specifying both a level of overall sequence identity to a member of the class and the presence of the correlated motif. Specifically, the claims specify that the DNA polymerase as an amino acid sequence that shows at least 30% overall identity with that of the polypeptide encoded by SEQ ID NO:4, and further includes a 15 amino-acid motif that is identical to one of SEQ ID NOs 5-22 except that it contains up to three (i.e., 0-3) amino acid substitutions as compared with the SEQ ID NO.

The claimed methods are described, inter alia, in original claims 9, 10; page 19, lines 19-20; page 31, lines 22-28; page 32, lines 1-3; and Table 3 on pages 20-21 of the specification. Support for claim 32 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; page 19, lines 19-20; page 31, lines 22-28; page 32, lines 1-3 and lines 10-16; and Table 3 on pages 20-21. Support for claim 33 is found in the specification as originally filed, inter alia, in original claim 10 and at page 19, lines 18-20. Support for claim 34 found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 35 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 36 is found in the specification as originally filed, inter alia, in

original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 37 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 38 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 39 is found in the specification as originally filed, inter alia, in original claim 19 Support for claim 40 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 41 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 42 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 43 is found in the specification as originally filed, inter alia, in original claims 13 and 18.

GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

The grounds of rejection to be reviewed on appeal are:

(1) Are claims 32-42 invalid for lack of written description under 35 U.S.C. § 112?

(2) Are claims 32 -42 invalid for lack of enablement under 35 U.S.C. § 112?

GROUPING OF CLAIMS

For reasons discussed below in the Argument section, the claims stand or fall together for purposes of ground of rejection numbered (1) above, as indicated below:

(1) Claims 32 and 39 stand or fall together.

(2) Claim 33 stands or falls alone.

(3) Claim 34 stands or falls alone.

(4) Claim 35 stands or falls alone.

(5) Claim 36 stands or falls alone.

(6) Claim 37 stands or falls alone.

(7) Claim 38 stands or falls alone.

(8) Claim 40 stands or falls alone.

(9) Claim 41 stands or falls alone.

(10) Claim 42 stands or falls alone.

For reasons discussed below in the Argument section, the claims stand or fall together for purposes of ground of rejection numbered (2) above, as indicated below:

(1) Claims 32 and 39 stand or fall together.

(2) Claim 33 stands or falls alone.

(3) Claim 34 stands or falls alone.

(4) Claim 35 stands or falls alone.

(5) Claim 36 stands or falls alone.

(6) Claim 37 stands or falls alone.

(7) Claim 38 stands or falls alone.

(8) Claim 40 stands or falls alone.

(9) Claim 41 stands or falls alone.

(10) Claim 42 stands or falls alone.

ARGUMENT

Ground of Rejection 1:

Claims 32 and 39 are not invalid for lack of written description

Pending claims 32-42 stand rejected for lack of written description. The Examiner states that claims 32-42 contain subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventors, at the time the application was filed, had possession of the claimed invention. This rejection is respectfully traversed. Reconsideration and withdrawal is requested.

The written description requirement serves both to satisfy the inventor's obligation to disclose the technologic knowledge upon which the patent is based, and to demonstrate that the patentee was in possession of the invention that is claimed. *Capon v. Eshhar*, 418 F.3d 1349, 1357 (Fed. Cir. 2005). To satisfy the written description requirement, the applicant does not have to utilize any particular form of disclosure to describe the subject matter claimed, but the description must clearly allow persons of ordinary skill in the art to recognize that he or she invented what is claimed. *Carnegie Mellon Univ. v. Hoffmann La Roche Inc.*, 541 F.3d 1115, 1122 (Fed. Cir. 2008) (quoting *In re Alton*, 76 F.3d 1168 (Fed. Cir. 1996)). In other words, the applicant must 'convey with reasonably clarity to those skilled in the art that, as of the filing date sought, he or she was in possession of the invention,' and demonstrate that by disclosure in the specification of the patent. *Id.* Such disclosure need not recite the claimed invention *in haec verba*, but it must do more than merely disclose that which would render the claimed invention obvious. *Univ. of Rochester v. G.D. Searle & Co.*, 358 F.3d 916, 923 (Fed. Cir. 2004). The descriptive text needed to meet the written description requirement varies with the nature and scope of the invention at issue, and with the scientific and technologic knowledge already in existence. *Capon*, 418 F.3d at 1357.

The present claims recite methods comprising steps of providing a particular type of DNA polymerase, contacting the DNA polymerase with a template, a primer that binds to the template, and a collection of nucleotides including at least one acyclonucleotide; and incubating the DNA polymerase with the template and the nucleotides so that the DNA polymerase extends

the primer by incorporating the nucleotides. Claim 32 specifies that the DNA polymerase has an amino acid sequence that shows at least 30% overall identity with that of the polypeptide encoded by SEQ ID NO:4, and further includes a 15 amino acid motif that is identical to one of SEQ ID NOs 5-22 except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO. The recited 15 amino acid motifs are shown in Table 3 of the specification at pages 20-21.

The Examiner has maintained that the written description requirement is not met for the scope of DNA polymerases encompassed by the claims. Appellants explain below that a structure/function relationship has been established for the DNA polymerases recited in the claimed methods, and that written description for the claims is more than satisfied under *Invitrogen Corp. v. Clontech Labs, Inc.*, 77 USPQ2d 1161 (Fed. Cir. 2005), and under the U.S. Patent and Trademark Office Written Description Training Materials (Revision 1, March 25, 2008).

A Structure/Function Relationship has been established.

The Examiner maintains the rejection for lack of written description on the ground that "applicants have not related the subgenus of structure to the acyclonucleotide incorporation function" (Office Action mailed April 21, 2009, page 4). Appellants respectfully disagree with this assertion. The disclosure of the specification, working examples, and declaratory evidence demonstrates a relationship between the structure recited in the claims and acyclonucleotide incorporation function. The claims require that the DNA polymerase have an amino acid sequence with at least 30% overall identity with that of the polypeptide encoded by SEQ ID NO:4 (Vent™). The claims also require that the DNA polymerase include a 15 amino acid motif that is identical to one of SEQ ID NOs 5-22, or has up to three amino acid substitutions.

The specification explains that proteins can display sequence similarity over short stretches of primary amino acid sequence (specification, page 14). These patches are thought to occur most often at essential protein interfaces, such as those involved in catalysis, substrate binding, or protein-protein recognition. The degree of sequence similarity, particularly in conserved sequence motifs, is predictive of the degree to which the proteins will behave

similarly in both physical properties and catalytic function (specification, page 14, lines 10-16). The claims include just such a motif, by requiring that the DNA polymerase include a 15 amino acid motif that is identical to one of SEQ ID NOs 5-22, or has up to three amino acid substitutions. The sequences of the 15 amino acid motifs and the DNA polymerase in which each is found are shown in Table 3 of the specification at pages 20-21. Each motif is within a conserved region having a role in substrate binding, known as "motif B" as defined by Delarue et al. (*Protein Eng.* 3:461-467, 1990; see citation to Delarue et al. in the specification at page 21 under Table 3; Delarue et al. was submitted with the Information Disclosure Statement filed on May 9, 2002, and is attached as **Exhibit A**). Delarue et al. does not recognize or discuss acyclonucleotide activity of any DNA polymerases. However, Delarue et al. indeed indicates that motif B is involved in DNA polymerase function. In the Discussion section, Delarue et al. states:

> ***From structure to function*. Considerable biochemical evidence points to the imporance of [motifs A, B and C] in the DNA polymerase activity**. A synthesized *E. coli* pol I oligopeptide corresponding to the N-terminal-most two-thirds of the loop region connecting helices O and P (**motif B**-see Figure 4) **has been shown to bind deoxynucleotide triphosphate substrates of pol I** as well as duplex DNA (Mildvan, 1989)" (Delarue et al., page 465, right col., lines 9-15; emphasis added).

This structure/function relationship between motif B and polymerase activity is further confirmed in declaratory evidence submitted during prosecution of the present application. In the Declaration by Dr. William Jack, filed on May 4, 2006 ("the Jack Declaration"; copy attached as **Exhibit B**), it states that Dr. Jack and colleagues have published articles in peer reviewed journals discussing the physical basis for the preferential incorporation of acyclonucleotides and enhanced incorporation with Vent A488L and 9°N 485L DNA polymerase mutants, citing to Gardner et al. (*J. Biol. Chem.* 279(12):11834-11842, 2004; Gardner et al. was submitted with the Jack Declaration and is attached as **Exhibit B**). Gardner et al. shows an alignment of Family B DNA polymerases in Figure 1. As is clear from the Figure, the "Region III" active site overlaps with the 15 amino acid motif recited in Appellants' claims. As provided in the Jack Declaration, Gardner discusses the physical basis for incorporation of acyclonucleotides at page 11841. This discussion mentions the A288 residue in Vent™, which is in the active site and in the 15 amino acid motif in Appellants' claims. A relationship between

"Region III", containing the 15 amino acid motif, and polymerase function, had been previously noted, e.g., in Hopfner et al. (*Proc. Nat. Acad. Sci. USA* 96:3600-3605, 1999; Hopfner et al. was submitted with the Jack Declaration and is attached as **Exhibit B**). Hopfner et al. reports the crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*. Hopfner et al. provide a structure based sequence alignment of archaeal family B polymerases, and show that Region III (which contains the 15 amino acid motif) is in the active site of these enzymes (see Hopfner et al., page 3603, col. 1, Figure 3, and col. 2, third full paragraph). Hopfner et al. discusses the conserved $KX_3NSXYGX_2G$ motif, which is a sub-motif within Appellants' claimed 15 amino acid motif, in the section entitled "Polymerase Active Site", noting that it and a second motif "form the bottom of the nucleotide-binding site" (Hopfner et al., page 3603, right col., lines 31-32). The subgenus of structure (i.e., the 15 amino acid motif) is clearly related to function.

Although there was recognition in the art that conserved motifs found in polymerases are involved in polymerase activity, it is Appellants who recognized and now claim a method of using a specific genus of polymerases which possess acyclonucleotide incorporation function. The set of 15 amino acid motifs specified by SEQ ID NOs 5-22 and recited in the claims are highly related to each other. SEQ ID Nos 6-17 differ from SEQ ID No 5 by three or fewer residues. SEQ ID Nos 18 and 20-22 differ from SEQ ID Nos 5 by six or fewer residues. Motifs of polymerases having sequences sharing less than 30% overall identity with Vent™ (and thus which are outside the scope of the claims) have motifs which differ from SEQ ID No 5 by seven or more residues (see, e.g., SEQ ID Nos 23-30 at page 21, Table 3 of the specification).

A structure/function relationship is not only supported by an understanding of the 15 amino acid motif and its role in enzymatic function. It is also supported by Appellants' working examples. Every DNA polymerase tested that meets the structural requirements of the claims has acyclonucleotide incorporation activity. Indeed, four different DNA polymerases, Vent™, Deep Vent™, *Pfu*, and 9N™, showed the ability to incorporate acyclonucleotides (specification, Example 6). Two variants of these enzymes, Vent™/A488L, and 9N™/A485L, were also shown in incorporate acyclonucleotides (specification, Example 11). By contrast, Thermosequenase, which is a Taq DNA polymerase variant that lacks the 15 amino acid motif required by the claims, showed a much stronger preference for dideoxyoligonucleotides over

acyclonucleotides (specification, Examples 5 and 12). The application therefore establishes the correlation between the sequence motif and the function recited in the claims.

In addition, the Jack Declaration includes an Appendix with data confirming that an archaeon Family B polymerase from *Methanococcus maripaludis*, having 41% sequence identity with Vent DNA polymerase, utilizes acyclonucleotides as a substrate (Jack Declaration Appendix I, attached as **Exhibit B**). Thus, support for a relationship between the DNA polymerases recited in the claims and acyclonucleotide incorporation function has been provided in by information in the specification regarding sequence similarity and function, exemplification of a relationship between the claimed structure and function in the specification, and data and information provided with the Jack Declaration.

The Examiner maintains his rejection without offering any reason *why* the claimed structure/function relationship allegedly has not been established. For example, in the Office Action mailed May 29, 2008, the Examiner said that "[w]hile Applicants comments regarding the homogeneity shared between this group of polymerases continues to be acknowledged, such is acknowledged in light of the degree of the vast majority of DNA polymerases, many of which have a high degree of homogeneity and not all of which share the ability to incorporate acyclonucleotides into a DNA fragment" (Office Action mailed May 29, 2008, page 4). Appellants have related specific structural features (overall sequence identity and the presence of a 15 amino acid motif in the active site of the enzyme) to function (acyclonucleotide incorporation function). The Examiner has provide no *reason* to doubt Appellants correlation. The Examiner is not entitled to substitute his personal skepticism for statements and evidence provided by the Appellants.

Written description support for the claims is met under Invitrogen Corp. v. Clontech Labs, Inc. 77 USPQ2d 1161 (Fed. Cir. 2005).

Relevant legal precedent also confirms that the written description requirement is satisfied for the present claims in view of the present specification. The decision in *Invitrogen Corp. v. Clontech Labs, Inc.* 77 USPQ2d 1161 (Fed. Cir. 2005) requires a finding that the claims

are adequately described. To emphasize this point, Appellants reiterate a close comparison between *Invitrogen* and the present claims here. The claim at issue in *Invitrogen* read:

> 1. An isolated polypeptide having DNA polymerase activity and substantially reduced RNAse H activity, wherein said polypeptide is encoded by a modified reverse transcriptase nucleotide sequence that encodes a modified amino acid sequence resulting in said polypeptide having substantially reduced RNase H activity, and wherein said nucleotide sequence is derived from an organism selected from the groups consisting of a retrovirus, yeast, Neurospora, Drosophila, primates and rodents.

The specification supporting the claim had only a single example of a polymerase having the recited activity. The court found that the claim met the written description requirement because, (1) at the time of the invention, sequences of reverse transcriptase (RT) genes were known; (2) members of the RT gene family shared significant homologies from one species to another; (3) the written description taught that the invention can be applied to RT genes of other retroviruses; and (4) the specification cited references providing the known nucleotide sequences of those genes.

It must be noted that, unlike the claim in *Invitrogen* which recites no structural limitations, the pending claims include explicit recitation of structural features (overall homology and a 15 amino acid motif). The present specification provides six specific examples of DNA polymerases that fall within the claims. As for the other factors from *Invitrogen*, (1) sequences of many DNA polymerases were known when the present application was filed; and (2) members of the DNA polymerase gene family share significant homologies from one species to another. See the present specification, e.g., at page 3, lines 8-21; and page 10, line 12, to page 15, line 34. For (3), the written description of the present case clearly teaches that the invention can be applied to DNA polymerases other than the ones specifically exemplified. See, for example, page 19, lines 15-27, which teaches:

> The similarity of incorporation patterns with these selected enzymes suggests that not only these archaeon DNA polymerases, but a larger family of DNA polymerases could share the ability to incorporate acyclo to a greater extent than dideoxy terminators. Since *Pfu*, Deep Vent® and 9°N™ DNA polymerases have greater than about 70% sequence identity with Vent DNA polymerase, other enzymes with equivalent or greater identity can reasonably be expected to perform as Vent® (exo-) DNA polymerase in this invention. Notably, those

> enzymes for which no significant sequence similarity is found (i.e., Family A DNA polymerases such as Taq) do not perform in similar ways in the current invention. This fact leads us to believe that archaeon enzymes showing intermediate identity, namely those between about 20 and 70% identity are reasonably candidates for this invention.

As to (4), the specification cites references providing the known sequences of such other DNA polymerases (see, for example, page 10, line 22; page 14, line 18; page 14, line 19; page 15, lines 19-24). Moreover, the sequences of other DNA polymerases are known and need not be fully presented in the specification to satisfy the written description requirement. See *Capon*, 418 F.3d at 1358.

Appellants maintain that, with regard to every relevant fact relied upon by the court, the present case has at least as much, or more description than was provided in *Invitrogen*.

The Examiner disputes this point because the claims encompass incorporation of acyclonucleotides into DNA and

> [t]his is not a property of a DNA polymerase that is well known in the art and the applicants have not adequately described this supposedly new function of a specific sub-genus of DNA polymerases. This is in contrast to the claims of Invitrogen in which the homologies of the encompassed DNA polymerases were high and that region responsible for reduced RNAse H activity in each of these DNA polymerases known such that the encompassed DNA polymerase variants known. (Office Action mailed April 21, 2009, page 5).

Appellants explain in detail the relationship between structure provided and acyclonucleotide function above. As discussed, Appellants have demonstrated (through several examples) that DNA polymerases that do have the claimed sequence do have the recited activities, and a DNA polymerase that does not have the claimed sequence does not have the recited activity.

Moreover, the fact that the present claims recite use to perform a newly discovered function (incorporation of acyclonucleotides) does not distinguish the present case from *Invitrogen*, as asserted by the Examiner. The claims in *Invitrogen* also related to DNA polymerases that have a new function (reduced RNAse H activity). The Examiner is correct that the *region* of DNA polymerase sequence that was responsible for RNAse H activity was previously known. As discussed above, the relevant region of DNA polymerases (region III)

involved in the present claims was also known (and known to be important for activity, just not for this activity). The present specification demonstrates that this known region is important for a new activity, much like the specification in *Invitrogen* demonstrated that changes in a known region could reduce activity. Closer factual scenerios in fact would be difficult to find!

Furthermore, Appellants fail to see how acyclonucleotide function of the DNA polymerases renders this case distinguishable from *Invitrogen*. In that case, a single example of an enzyme having a desired function (reduced RNAse H activity) was adequate to support the claims.

Appellants' recognition of a class of polymerases which incorporate acyclonucleotides is new, and Appellants have linked the functional activity with structure and a characterized structural, functional motif (i.e., the 15 amino acid motif). There is no basis for distinguishing the present case from *Invitrogen*. The Examiner suggests that *Invitrogen* is not applicable because "the homologies of the encompassed DNA polymerases were high." Yet the *Invitrogen* claim is completely devoid of structural limitations, and recites polymerases from organisms as diverse as viruses, yeasts, and primates! If unspecified sequences from such varied species have "high" homology in the Examiner's view, Appellants fail to understand how homologies between sequences encompassed by the present claims, which recite concrete structural limitations, are not also "high."

In a further attempt to distinguish *Invitrogen*, the Examiner stated that

> the description held by Invitrogen is specific to the claims of invitrogen [sic], based upon the specification and art as well as a. Actual reduction to practice, b. Disclosure of drawings or structural chemical formulas, c. Sufficient relevant identifying characteristics, such as: Complete structure, ii. Partial structure, iii. Physical and/or chemical properties, iv. Functional characteristics when coupled with a known or disclosed correlation between function and structure, d. Method of making the claimed invention, e. Level of skill and knowledge in the art and f. Predictability in the art. (Office Action mailed April 21, 2009, carryover paragraph from pages 5-6).

Legal decisions would be meaningless as precedent if they could be applied only to a single set of facts. Appellants have provided a close comparison of (i) the facts in the *Invitrogen* case and the (stronger) facts here; and (ii) the claims of the present application and a claim from

Invitrogen for which written description was affirmed. There has been no showing that *Invitrogen's* claimed genus all had "high" homology or "known" function such that the present claims can be distinguished from the case. No other bases for finding *Invitrogen* inapplicable have been offered.

Written description support for the claims is met under the U.S. Patent and Trademark Office Written Description Training Materials

The Examiner compared the present claims to the U.S. Patent and Trademark Office Written Description Training Materials (hereinafter, the "Guidelines") and found lack of description in Appellants' claims compared to claim 2 in Example 11 of the Guidelines because "claim 2 is drawn to a nucleic acid having 85% identity to a specific sequence, a partial structure. This is relative to the instant claims which require even less partial structure of 30% identity." (Office Action mailed April 21, 2009, page 7).

The Examiner has not analyzed Appellants' claims in view of the knowledge of DNA polymerase structure and the requirement of a conserved motif which is associated with enzymatic function. Claim 2 in Example 11 of the Guidelines concerns a claim to nucleic acid encoding hypothetical polypeptide having "activity X". In contrast to the present claims, the hypothetical polypeptide encoded by the nucleic acid does not share significant sequence identity with any known polypeptide or polypeptide family. Also unlike the present claims, the specification for this hypothetical example discloses only a single nucleic acid sequence that encodes a polypeptide having "activity X". Any comparison of the present claims to Example 11 should take these facts into consideration. Another important factor for analysis in Example 11 is the presence of a disclosed or art-recognized correlation between structure and function. Appellants have provided this correlation.

Example 5 of the Guidelines presents a fact pattern much more analogous to Appellants claims, and is a more appropriate basis for comparison. Example 5 concerns a claim to an "isolated protein comprising Protein A," wherein Protein A includes the amino acid sequence of SEQ ID NO:1, has the ability to bind and activate Protein X, and is purified by a recited set of conditions. The sequence of SEQ ID NO:1 in this hypothetical claim has 10 amino acids.

Likewise, Appellants' claims recite DNA polymerases that include a 15 amino acid motif and have a specific binding and activity function, which is the ability to incorporate acyclonucleotides in a polymerase extension reaction. The polymerases are not defined by purification conditions. However, significant structural definition for the polymerases is provided by requiring at least 30% identity to SEQ ID NO:4.

In the hypothetical fact pattern set forth for Example 5, claim 1, the specification fails to disclose the complete structure of Protein A and it fails to disclose any art recognized correlation between the structure of the claimed protein and its function of binding and activating Protein X. Nonetheless, written description is affirmed for the claim because the specification discloses a partial (10 amino acid) sequence of Protein A and because relevant identifying characteristics are provided in the form of its ability to bind and activate Protein X, and purification features.

If anything, the present specification provides more description support for the claims than is provided for claim 1 of Example 5 of the Guidelines. Appellants' specification describes examples of complete structures for polymerases that fall within the claims. Appellants' 15 amino acid motif imposes greater structural definition for a polymerase than the 10 amino acid sequence defining the hypothetical polypeptide of Example 5. Appellants' polymerases possess a binding ability and activity (acyclonucleotide incorporation) which is just as well defined as those of the hypothetical polypeptide of Example 5. Whereas no correlation of protein structure with function is provided in Example 5, Appellants' provide detailed structure/function correlation, as set forth above. In this aspect, Appellants provide more support than the Guidelines require. Another factor favoring support for the hypothetical polypeptide was the specification's disclosure of methods for isolating the polypeptide and a working example showing the polypeptide was successfully isolated. Appellants' have also shown that one of skill in the art can make and use polypeptides as claimed, and that polypeptides have the recited function.

Claim 33 is not invalid for lack of written description

Claim 33 stands rejected for lack of written description. Claim 33 specifies that the DNA polymerase has an amino acid sequence that shows at least 70% overall identity with that of SEQ

ID NO:4. Because this claim requires a higher overall identity to SEQ ID NO:4, the genus of polymerases encompassed by the claim is smaller than that of claim 32. Thus, the level of description required is reduced as compared with claim 32. Appellants' specification demonstrates that multiple polymerases within the genus possess acyclonucleotide function. (Appellants emphasize that polymerases from the broader genus have this function as well; see the Jack Declaration, Appendix I, which shows that a *Methanococcus* DNA polymerase having only 41% sequence identity to Vent™ DNA polymerase incorporates acyclonucleotides more efficiently than dideoxynucleotides.) Even if claim 32 were not fully supported by the specification (which Appellants do not concede), claim 33 would be.

Claim 34 is not invalid for lack of written description

Claim 34 stands rejected for lack of written description. Claim 34 specifies that the 15 amino acid motif is identical to one of SEQ ID Nos 5-22. Given the further limitation on the sequence of the motif (i.e., such that the motif does not include amino acid substitutions), the genus of polymerases encompassed by the claim is smaller than that of claim 33. The level of description required for this claim is reduced as compared with claim 32. Even if claim 32 were not fully supported by the specification, claim 34 would be.

Claim 35 is not invalid for lack of written description

Claim 35 stands rejected for lack of written description. Claim 35 specifies that the 15 amino acid motif is identical to one of SEQ ID NOs 15-17, except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO. Because it covers fewer motifs, this claim refers to a genus of polymerases that is smaller than that encompassed by claim 32. The level of description required to support this claim is less than required for claim 32.

Claim 36 is not invalid for lack of written description

Claim 36 stands rejected for lack of written description. Claim 36 specifies that the 15 amino acid motif is identical to one of SEQ ID Nos 5-17. The genus of polymerases

encompassed by this claim is even smaller than that of claim 32 and requires less description to be adequately supported.

Claim 37 is not invalid for lack of written description

Claim 37 stands rejected for lack of written description. Claim 37 specifies that the 15 amino acid motif is identical to one of SEQ ID NOs 5-8 except that it may contain up to three amino acid substitutions. Again, the genus of polymerases encompassed by this claim is even smaller than that of claim 32 due to further limitation of the 15 amino acid motif and is fully supported by the specification.

Claim 38 is not invalid for lack of written description

Claim 38 stands rejected for lack of written description. Claim 38 specifies that the amino acid motif is identical to one of SEQ ID NOs 5-8. The genus of polymerases encompassed by this claim is smaller than that of claim 32 due to further limitation of the 15 amino acid motif and is fully supported by the specification.

Claim 40 is not invalid for lack of written description

Claim 40 stands rejected for lack of written description. Claim 40 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-22. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is fully supported by the specification.

Claim 41 is not invalid for lack of written description

Claim 41 stands rejected for lack of written description. Claim 41 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-17. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is fully supported by the specification.

Claim 42 is not invalid for lack of written description

Claim 42 stands rejected for lack of written description. Claim 42 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID Nos 5-8. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is fully supported by the specification.

In conclusion, the provided teachings in the specification, examples, sequences, declaratory evidence, and data are more than sufficient to describe function and support description of the claims. The Examiner has not established otherwise. For reasons set forth above, withdrawal of the rejection of claims 32 and 39 as allegedly lacking written description is respectfully requested.

Ground of Rejection 2:

Claims 32 and 39 are not invalid for lack of enablement

Pending claims 32 and 39 stand rejected for lack of enablement. The Examiner states that the specification, while being enabling for a method comprising providing a DNA polymerase selected from the group consisting of Vent™, Deep Vent™, *Pfu*, and 9°TM or the specifically disclosed variants of claim 43, "does not reasonably provide enablement for any method comprising providing a DNA Polymerase having an amino acid sequence that shows a mere 30% overall identity with that of SEQ ID NO:4 and further includes a 15 amino-acid motif that is identical to SEQ ID NO:5 except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO..." (Office Action mailed April 21, 2009, pages 8-9). The Examiner stated that "determination of those DNA polymerases having the desired biological characteristics is unpredictable and the experimentation left to those skilled in the art is unnecessarily, and improperly, extensive and undue" (Final Office Action mailed April 21, 2009, pages 11). Appellants have previously reviewed the factors set forth in *In re Wands* (858 F.2d 731, 8 USPQ2nd 1400, Fed. Cir. 1988) with respect to the present claims and review them here, in response to the Examiner's assertion that Appellants' burden has not been met.

First, Appellants address the Examiner's comments regarding *Wands* factor (2), Amount of Direction or Guidance. In the Final Office Action mailed April 21, 2009, the Examiner maintained that guidance was lacking as to DNA polymerases which have the ability to incorporate acyclonucleotides into a DNA template, and requested clarification as to how the 15 amino acid motif correlates with acyclonucleotide function (Office Action mailed April 21, 2009, page 12). As explained above in the arguments for written description, the 15 amino acid motif is a highly conserved motif in the active site of family B DNA polymerases which plays a role in substrate binding. The Examiner disputes a structure/function correlation because "applicants have not disclosed such a single motif but rather continue to refer to any of a number of motifs or variants thereof"(Office Action mailed April 21, 2009, page 12). Some variability within the genus of motifs is permitted, given that variable polymerases share acyclonucleotide incorporation function. For example, both 9°N polymerase and Vent™ incorporate

acyclonucleotides, although their 15 amino acid motifs differ by three amino acids (compare SEQ ID NO 5 and SEQ ID NO 7 at page 20, Table 3 of the specification). A *Methanococcus maripaludis* DNA polymerase having a more divergent sequence also possesses acyclonucleotide incorporation activity. The claims do require a degree of conservation of sequence, which is clearly expressed in the claims. The fact that variable polymerases share a specific function does not render them "unpredictable."

As to the (1) Quantity of Experimentation Necessary, and (3) Presence or Absence of Working Examples, Appellants reiterate that one or ordinary skill could make and test all polypeptides within the scope of the claims to determine their ability to extend a DNA primer or incorporate acyclonucleotides (including to determine their ability to preferentially select acyclonucleotides). Appellants' working examples include demonstration of activity of multiple species of DNA polymerases set forth in the specification and in declaratory evidence discussed herein.

As to (5) State of the Prior Art, and (7) Predictability of the Art, Appellants note, and the Examiner has acknowledged, that the prior art with regard to DNA polymerases and their classification is extensive. However, Appellants disagree with the Examiner's assertion that "determination of those DNA polymerases having the desired biological characteristics is unpredictable and the experimentation left to those skilled in the art is unnecessarily, and improperly, extensive and undue" (Office Action mailed April 21, 2009, page 11). Appellants have identified polymerases which have acyclonucleotide incorporation function by virtue of structural and physical characteristics distinctive of well-characterized DNA polymerases. These characteristics include overall sequence identity to a polymerase, and the presence of a conserved motif. Appellants have shown that all members tested within the genus of polymerases have the recited activity. The Examiner's only discernible reason for declaring these features unpredictable is the breadth of the genus of polymerases. This improperly disregards Appellants' demonstration of activity for multiple species. It also disregards the fit of Appellants' observed activity with well characterized classification schemes for DNA polymerases (among which one finds substantial variability despite conserved nucleotide polymerase activity).

As to the (4) Nature of the Invention and (8) Breadth of the Claims, Appellants reiterate that DNA extension reactions are well within the skill of those of ordinary skill. As part of their invention, Appellants have described a class of DNA polymerases that can incorporate acyclonucleotides, and have shown function for six different species within the class. Given the demonstrated correlation of structure with function and other reasons provided above, Appellants disagree with the Examiner's assertions that the scope of the claims is not enabled.

As to (6) Relative Skill of those in the Art, Appellants submit, and the Examiner has agreed, that the relative skill of those in the art is very high.

Claim 33 is not invalid for lack of enablement

Claim 33 stands rejected for lack of enablement. Claim 33 depends from claim 32 and specifies that the DNA polymerase has an amino acid sequence that shows at least 70% overall identity with that of SEQ ID NO:4. Because this claim requires a higher overall identity to SEQ ID NO:4, the breadth of the claim is smaller than that of claim 32. The scope of enablement provided by the disclosure is more than sufficient to support the scope of this claim, not least because multiple polymerases that fall within the claimed genus are exemplified.

Claim 34 is not invalid for lack of enablement

Claim 34 stands rejected for lack of enablement. Claim 34 depends from claim 32 or 33 and specifies that the 15 amino acid motif is identical to one of SEQ ID Nos 5-22. This further limitation on the sequence of the motif (i.e., such that the motif does not include amino acid substitutions) provides a claim of smaller breadth than claim 32 and which is more than supported by the disclosure.

Claim 35 is not invalid for lack of enablement

Claim 35 stands rejected for lack of enablement. Claim 35 depends from claim 32 or 33 and specifies that the 15 amino acid motif is identical to one of SEQ ID NOs 15-17, except that it

contains up to 3 amino acid substitutions as compared with the SEQ ID NO. This claim covers fewer motifs than claim 32 and is enabled for its full scope.

Claim 36 is not invalid for lack of enablement

Claim 36 stands rejected for lack of enablement. Claim 36 specifies that the 15 amino acid motif is identical to one of SEQ ID NOs 5-17. Again, the genus of polymerases encompassed by this claim is even smaller than that of claim 32 and is enabled by the disclosure provided.

Claim 37 is not invalid for lack of enablement

Claim 37 stands rejected for lack of enablement. Claim 37 specifies that the 15 amino acid motif is identical to one of SEQ ID NOs 5-8 except that it may contain up to three amino acid substitutions. The genus of polymerases encompassed by this claim is even smaller than that of claim 32 due to further limitation of the 15 amino acid motif and is fully enabled by the specification.

Claim 38 is not invalid for lack of enablement

Claim 38 stands rejected for lack of enablement. Claim 38 specifies that the amino acid motif is identical to one of SEQ ID NOs 5-8. The genus of polymerases encompassed by this claim is smaller than that of claim 32 due to further limitation of the 15 amino acid motif and is fully enabled by the specification.

Claim 40 is not invalid for lack of enablement

Claim 40 stands rejected for lack of enablement. Claim 40 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-22. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is fully enabled by the specification.

Claim 41 is not invalid for lack of enablement

Claim 41 stands rejected for lack of enablement. Claim 41 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-17. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is fully enabled by the specification.

Claim 42 is not invalid for lack of enablement

Claim 42 stands rejected for lack of enablement. Claim 42 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID Nos 5-8. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is enabled for its full scope.

In light of the above, Appellants submit that claims 32 -42 satisfy the enablement requirement. Allowance of the claims is requested.

Respectfully submitted,

Date: December 23, 2009

/Margo H. Furman/
Margo H. Furman, PhD, JD
Registration Number 59,812
Tel. No. (617) 248-4073

PATENT GROUP
CHOATE, HALL & STEWART LLP
Two International Place
Boston, MA 02110
Tel: (617) 248-5000
Fax: (617) 502-5002
patentdocket@choate.com

CLAIMS APPENDIX

1-31. (Canceled)

32. (Previously presented) A method comprising steps of:

providing a DNA polymerase having an amino acid sequence that shows at least 30% overall identity with that of the polypeptide encoded by SEQ ID NO:4, and further includes a 15 amino-acid motif that is identical to one of SEQ ID NOs 5-22 except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO;

contacting the DNA polymerase with a template, a primer that binds to the template, and a collection of nucleotides including at least one acyclonucleotide; and

incubating the DNA polymerase with the template and the nucleotides so that the DNA polymerase extends the primer by incorporating the nucleotides.

33. (Previously presented) The method of claim 32, wherein the DNA polymerase has an amino acid sequence that shows at least 70% overall identity with that of SEQ ID NO:4.

34. (Previously presented) The method of claim 32 or claim 33, wherein the 15 amino-acid motif is identical to one of SEQ ID NOs 5-22.

35. (Previously presented) The method of claim 32 or claim 33, wherein the 15 amino-acid motif is identical to one of SEQ ID NOs 5-17 except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO.

36. (Previously presented) The method of claim 35, wherein the 15 amino acid motif is identical to one of SEQ ID NOs 5-17.

37. (Previously presented) The method of claim 32 or 33, wherein the 15 amino acid motif is identical to one of SEQ ID NOs 5-8 except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO.

38. (Previously presented) The method of claim 37, wherein the 15 amino acid motif is identical to one of SEQ ID NOs 5-8.

39. (Previously presented) The method of claim 32 or 33, wherein the step of incubating comprises incubating the DNA polymerase with the template and the nucleotides so that the DNA polymerase extends the primer by incorporating the nucleotides, and preferentially incorporates acyclonucleotides.

40. (Previously presented) The method of claim 32 or 33, wherein the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-22.

41. (Previously presented) The method of claim 35, wherein the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-17.

42. (Previously presented) The method of claim 37, wherein the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-8.

43. (Previously presented) The method of claim 32 or 33 wherein the DNA polymerase is Vent™, Deep Vent™, 9°N, *Pfu*, Vent™/488L, or 9°N/485L.

EVIDENCE APPENDIX

Appellants had provided the following evidence during prosecution of the instant application:

Exhibit A: Delarue et al., *Protein Eng.* 3:461-467, 1990. This reference was cited in the Information Disclosure Statement and Form PTO-1449 filed on May 9, 2002, and was entered into the record on May 13, 2002. The Form PTO-1449 was initialed by the Examiner on September 29, 2004, confirming that the reference was entered into the record.

Delarue et al. is attached hereto at pages 33-39.

Exhibit B: Declaration of William Jack, accompanying references and Appendix I. The Declaration was submitted with four references, listed below, and Appendix I along with a response to Office Action filed May 4, 2006, and was entered into the record in PAIR on May 9, 2006 as the entry designated "Rule 130, 131 or 132 Affidavits." Entrance into the record was confirmed by the Examiner's reference to this Declaration on page 3 of the Advisory Action mailed on July 5, 2006.

The Declaration of William Jack is attached hereto at pages 40-47.

Rodriguez et al., *J. Mol. Biol.* 299:447-462, 2000, is attached hereto at pages 48-63.

Gardner et al., *J. Biol. Chem.* 279(12): 11834-11842, 2004, is attached hereto at pages 64-72.

Hashimoto et al., *J. Mol. Biol.* 306:469-477, 2001, is attached hereto at pages 73-81.

Zhao et al., *Structure* 7(10):1189-1199, 1999, is attached hereto at pages 82-92.

Hopfner et al., *Proc. Nat. Acad. Sci. USA* 96:3600-3605, 1999, is attached hereto at pages 93-98.

Appendix I is attached hereto at pages 99-101.

RELATED PROCEEDINGS APPENDIX

Not applicable.

EXHIBIT A

Protein Engineering vol.3 no.6 pp.483–487, 1990

An attempt to unify the structure of polymerases

Marc Delarue, Olivier Poch†, Noel Tordo†, Dino Moras and Patrick Argos‡

Laboratoire de Cristallographie and †Biochimie, IBMC du CNRS, 15 rue René Descartes, 67084 Strasbourg, ‡Protein Engineering Department, Rhône-Poulenc Rorer, 20 Avenue Raymond Aron, 92165 Fontenay-sous-Bois, †Unité de Virologie Moléculaire, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, ‡Institut für Molekulare Biologie Laboratory, Postfach 10.2209, Meerbuschstrasse 1, 6900 Heidelberg, FRG

With the great availability of sequences from RNA- and DNA-dependent RNA and DNA polymerases, it has become possible to delineate a few highly conserved regions for various polymerase types. In this work a RNA polymerase sequence from bacteriophage SP02 was found to be homologous to the polymerase domain of the Klenow fragment of polymerase I from *Escherichia coli*, which is known to be closely related to those from *Staphylococcus pneumoniae*, *Thermus aquaticus* and bacteriophages T7 and T5. The alignment of the SP02 polymerase with the other five sequences satisfactorily recovered the conserved motifs in these proteins. Three of the motifs matched reasonably all the conserved motifs of another DNA polymerase type, characterized by human polymerase α. It is also possible to find these three motifs in mammalian DNA-dependent RNA polymerases and two of them in RNA polymerases I and RNA terminal transferases. These latter two motifs also matches two of the four motifs recently identified in all RNA-dependent polymerases. From the known tertiary architecture of the Klenow fragment of *E. coli* pol I, a spatial arrangement can be inferred for these motifs. In addition, numerous biochemical experiments suggesting a role for the motifs in a common functions (dNTP binding) also support these inferences. This speculative hypothesis, attempting to unify polymerase structure at least locally, if not globally, under the pol I fold, should provide a useful model to direct mutagenesis experiments to probe template and substrate specificity in polymerases.

Key words: catalytic domain/DNA polymerases/RNA polymerases/sequence/structure

Introduction

The number of available protein sequences is growing rapidly due to the facility of nucleotide sequencing techniques. One particular class evolved and augmented as the polymerase family which is central to the duplication and expression of genes. Polymerases can use RNA or DNA as a template (RNA- or DNA-dependent); the product can also be RNA or DNA. Polymerases are found both in eukaryotes and prokaryotes, though sequencing efforts have often concentrated on those from viruses. One way to use the information contained in all these sequences is to try to align them and thereby discover their strongest internal features and similarities. This has been achieved for DNA-dependent DNA polymerases, where three main subfamilies have been identified. One of them contains the Klenow fragment of *Escherichia coli* polymerase I, whose three-dimensional structure is known (Ollis *et al.*, 1985b), and polymerases from phages T5 (Ollis *et al.*, 1985b; Argos *et al.*, 1986) and T3 (Hanvit and Ito, 1989), and from *Thermus aquaticus* (Lawyer *et al.*, 1989) and *Staphylococcus pneumoniae* (Lopez *et al.*, 1989). This family will be referred to as the pol I family. For another set of DNA-dependent DNA polymerases (Wong *et al.*, 1988), of those homologous to the human polymerase α (hereafter referred to as pol α), more than 10 sequences from various species are known. A third subfamily of DNA-dependent DNA polymerases, homologs called the pol B type, has only two members: DNA polymerase II (Matsukage *et al.*, 1987) and terminal transferase (Peterson *et al.*, 1985; Kotani, 1986; Zmudzka *et al.*, 1988). Until now, just one DNA-dependent DNA polymerase sequence, from the SPO1 bacteriophage (Hahn and Rabussay, 1988; Jong *et al.*, 1989), resisted alignment with any of the three aforementioned types.

Clearly, the aim of these alignments, apart from evolutionary implications, is the identification of the regions essential for polymerase function, since these sequence segments should appear in the most conserved. Generally, a good number of aligned sequences will ensure sufficient variability to identify the functionally required regions. For the pol I type, the five previously aligned (Ollis *et al.*, 1985b; Argos *et al.*, 1986; Lawyer and Ito, 1989; Lopez *et al.*, 1989) sequences are sufficiently close so as not to allow confident delineation of the absolutely required motifs.

In the present work, it has been found that the polymerase from bacteriophage SPO2 can be aligned, using a sensitive method, with the polymerase portion of the Klenow fragment in the C-terminal part of the protein [the N-terminal domain has a 3′-5′ exonuclease function (Freemont *et al.*, 1988)]. The total alignments of the C-terminal part of the six proteins of the pol I type is presented. The relationship of SPO2 polymerase with those from phages T7 and T5, *Synechococcus*, Zmudzka and Kotani is sufficiently distant that highly conserved regions can now be referred to five to another. Interestingly, three of the five regions match reasonably with the three most conserved motifs of DNA-dependent DNA pol α, suggesting that the two polymerase types may share a common tertiary fold, or at least common shorter local tertiary architecture required for similar functions. These motifs are likely to represent modules required for the polymerase structure and activity.

Searches were then performed to detect such sequence patterns in other polymerase families. All three motifs could be found in DNA dependent RNA polymerases that consist of only one subunit (see Mosorin *et al.*, 1987); two motifs were found in pol β, in the same linear arrangement and maintaining the strictly conserved residues. In addition to this, an examination of several aligned RNA-dependent RNA polymerases as well as reverse transcriptases, for which four highly conserved motifs have been highlighted (Kamer and Argos, 1984; Poch *et al.*, 1989), all lent the suggestion that two of these motifs could match the two motifs shared by RNA pol β, pol α and pol β. These sequence similarities are further supported by a consistently significant alignment between the entire polymerase domain of two members of these different families, namely a DNA-dependent RNA

Fig. 1. Homology search curve (open, 1987) between the Klenow fragment of DNA pol I from *E.coli* and DNA polymerase from bacteriophage RB49. The scores, whose lengths ranged from 3 or 20 to more than 200, for search peaks are plotted over the entire relative length, with the largest score designating either overlap errors from the matching windows. The peak values (D) are scored as a measure of standard deviations above the statistic mean for each relative length. Thus these values are 6 to 8, 3 to 5, and 2 to 3 for short, medium and long lengths, respectively. The curve is kept to affect the alignments. If Figure 2 is indicated by arrows. The remaining alignments from fit Figure 2 are found on supplementary pages 1 of all the sequences.

polymerases and an RNA-dependent DNA polymerase. The alignment of the C-terminus core motif, containing two acidic groups, is in agreement with those suggested by Argos (1988b) amongst RNA-dependent RNA polymerases, reverse transcriptases and DNA-dependent DNA pol αα.

An examination of these conserved regions relative to the known Klenow fragment ternary enzyme/core positions them in cross-links to interact with nucleotides. In fact, many biochemical data (Joyce and Steitz, 1987; Lecler *et al.*, 1987b) point to the importance of these motifs in the catalytic process, in the different families of polymerases. The design of site-directed mutagenesis experiments should benefit from the model suggested here, replacing possible common tertiary segments for various polymerases, if not a common topology.

Results

Pairwise sequence comparisons were done by a search procedure based on similar characteristics (Argos, 1987). The resulting alignments of the C-terminal polymerase portion of the Klenow fragment of DNA-dependent DNA polymerase I from *E.coli* and polymerases from bacteriophages T7, T5, *S.pneumoniae* and *S.aureus* are similar to the ones given by Argos *et al.* (1986a), Lewis and Jen (1989), Lopez *et al.* (1988) and Lawyer *et al.* (1989) respectively. The DNA polymerase sequence from bacteriophage SPO2 was also compared to the *E.coli*, *S.pneumoniae*, T3 and T7 polymerases. The search matrix and

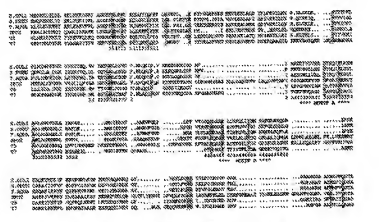


Fig. 2. Multiple sequence alignment of the C-terminus regions of *E.coli* DNA polymerase I, *S.pneumoniae*, *T.aquaticus*, bacteriophages SPO2, T5 and T7 polymerases. The major conserved motifs are numbered 1–3. The number of the first amino acid of the displayed sequence in T5, 511, 405, 393, 552 and 565 for *E.coli*, *S.pneumoniae*, *T.aquaticus*, SPO2, T3 and T7 respectively.

alignment pattern where the SPO2 and E. coli polymerases are shown in Figure 1; it is clear that strong regions of homology exist at the 4.5 SD or higher level. No such strong relationships could be found between SPO2 and S. pneumoniae, T. aquaticus, T7 or T5 polymerases. Figure 2 shows the alignment of the C-terminal part of these six polymerase sequences; this multiple alignment was obtained by manual adjustment of the different but closest pairwise alignments. A conservation profile resulting from this alignment was also calculated (data not shown); this profile is based on a five-residue window and the score is simply the normalized sum of the matrix elements corresponding to the residues observed in all the different pairwise alignments. The

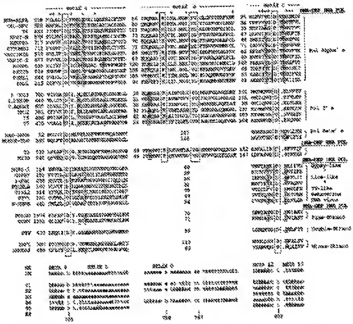


Fig. 3. Multiple alignment of various polymerase sequences for motifs A, B and C. [illegible]

USSN 10/089,027

Atty. Docket No. 2007651-0001

4567604v1

motifs in a template-DependE matrix as used by Gribskov and Burgess (1986). A threshold of 60% conservation hides five regions of high homology, which are indicated in Figure 2. These regions are characterized by the consensual sequences (F)(R>=+)=(=Q)(L)=***, in the first region (* refers to a given alignment position occupied by any amino acid; () indicates an option and shared unequivocally conserved by the six sequences, and a single space refers to exact conservation in all the sequences), (T)(Q)F for region 2, D=(=)=*=L for region 3, K=====Y(Q) for region 4 and VH(E>)G for region 5.

A multiple alignment was also built for 12 sequences of DNA-dependent DNA polymerases on three main, yeast and several viruses. These sequences are essentially the ones generated by Wang et al. (1989) together with two recent additions to the nrBBP sequence data bank: toxic polymerase from another adenovirus, Wrair-G, and one from a nuclear cytoplasmic virus, Iridoviprivus. The resulting conservation profile using a window of five residues left three regions above a 60% conservation threshold. They are characterized by D=*=(M)-(Y)(F)(L), K=*=S(K>)(T)G and (Y)(Q)(L)DE, as previously noted by other authors (Bernad et al., 1987). A comparison of the most conserved sequences for the pol I and pol α types suggests an overlap of these three regions (I and pol α with motifs 3, 4 and 5 of the pol I type; these regions will hereafter be called motifs A, B and C (Figure 3). These three motifs are common on aromatic residues motif A contains a strictly conserved aspartate at the junction of a beta strand and an alpha helix, motif B contains an alpha helix with positive charges and motif C has a district of negative charges location in a beta-turn-beta secondary structure, as implied from the Klenow structure (Figure 4). Great variability in the lengths separating the different conserved regions was observed for the pol α: 90–120 residues between motifs A and B (average 98) but 25–150 residues between motifs B and C (average 68). In the pol β, these values correspond respectively with 30 and 100 residues, on average.

A set DNA-dependent DNA polymerase β (Mossenbaugh et al., 1987) can be easily aligned with a known tertiary structure, the bacteriophage T4 enzyme (Wang et al., 1989). Though many residues are identically conserved amongst the two sequences, as N-terminal and C-terminal fragments may be related to motifs A and C of DNA polymerases (Figure 3).

For DNA-dependent RNA polymerases, the three motifs could only be found in the putative units of only one subunit. Four different sequences of this type are known, the two most distantly related being from phage T3 and yeast mitochondria (Masters et al., 1987). The distances between motifs computed well with the ones of the pol I type. The sequence fragments corresponding to the motifs are presented in Figure 3, together with those found in DNA, pol α, pol β and pol β'.

Secondary structure predictions averaged over each set of sequences (DNA pol, α, pol I, and pol β; and DNA-dependent RNA pols) were made according to the procedures of Argos (1985) and are given in Figure 3. Except for pol β, for which only two sequences are known, they often show agreement amongst themselves as well as with the pol I structure, especially for motifs A and C.

Recently, Poch et al. (1989) identified four motifs shared by all RNA-dependent RNA and DNA polymerases (84 different sequences). Two of these motifs were found to be similar to two of those contained in the DNA-dependent polymerases, namely, motifs A and C. These two motifs contain the strictly conserved residues of the same name mentioned. The tertiary structure predictions agree with the observed structure in the corresponding

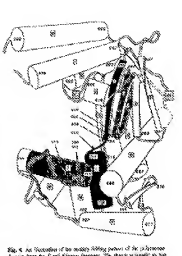


Fig. 4. An illustration of the motifs sitting pattern of the β-hairpin domain from the *E.coli* Klenow fragment. The thick α-helix is not given in Ollis et al. (1988). The four α-helix domain regions, in ascending from the N-terminus to the C-terminus, correspond respectively to motifs A (dark region), B (narrow dotted) and C (horizontal stripes), whose sequences are given in Figure 3. Sequence positions relative to the secondary structures are mentioned. Asterisks are conserved unanimously and α-helices are designated by letters according to the alphabet. The structural position of the conserved lysine acids is indicated in each of the motifs

regions of the Klenow fragment (Ollis et al., 1988). Sequences fragments from 12 reconstructive RNA-dependent polymerases (1.1 from each of the subfamilies: reverse transcriptases, plant-viruses and double-strand RNA polymerases) are shown aligned in three matched in the DNA pols (Figure 3).

Several total RNA-dependent DNA polymerases and reverse transcriptase sequences were then compared to those from pol I and pol α using the same sequence comparison techniques as above (Argos, 1987). Although the different polymerase comparisons, extensive sequence similarities were found between the total hepatitis B RNA-dependent DNA polymerase from Woodchuck (Gellbert et al., 1987) and Herpes simplex virus DNA polymerase, a member of the DNA pol α family (Gibbs et al., 1985). The motifs as shown in Figure 5 and the resultant alignment in Figure 6. It is noteworthy that 31% of the aligned residues are identical. In this alignment, two conserved regions can be delineated that correspond to motifs A and C. The homology between these two segments provides a possible link between the DNA- and RNA-dependent polymerases.

Discussion

DNA-dependent DNA polymerases

Sequence similarities. In this work, a new member was added to the pol I type of DNA-dependent DNA polymerases. An analysis of the multiple alignments for the pol I types as well as for the pol α type pointed to three conserved motifs, built as

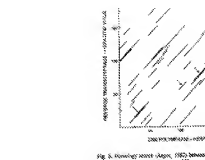


Fig. 3. Homology search (Argos, 1987) between bacteriophage consensus DNA polymerase and Herpes simplex DNA polymerase. The search window length ranged from 9 to 30 in steps of 3; the search peaks are plotted over the same window length, with the longest value corresponding to the most exacting search from the histograms. The peak values are all scored as a number of standard deviations (σ) above the mean score for each window length. They have indexes I for $\sigma \geq 3.0$; good fits, $4.0 \leq I < 5.0$; excellent fits, $I \geq 5.0$. The best peak is seen to start in region A as shown in Figure 4 (a reference by arrow).

sponsible. The distances between these motifs are quite variable. However, within each polymerase type, there is already great conservation in the distances between conserved regions. Moreover, it would seem unlikely that those of the most conserved sequence regions in the polymerase domains of pol I-type (Figure 2) could be aligned with the three most conserved regions of DNA pol α purely by chance. A different piece of evidence pointing to a possible link between the pol I and the pol α types comes from sequence similarities within the extracellular domain of the T5 polyvirion and pol III (Leavitt and Ito, 1988). It is thus reasonable to expect that polymerase domains to be related also. In fact, this observation has been recently extended to all extracellular domains of pol III and pol II, including SPO2 (Bernad *et al.*, 1989). In addition, the secondary structure predictions, especially for motifs A and C, point to sequence-induced structures for pol III that are similar to those observed in the pol I Klenow structure.

From the sequences to structure. The three motifs A, B and C are clustered and conservative in the C-terminal fraction of the Klenow fragment and correspond to structural features likely to interact with the DNA (Ollis *et al.*, 1985). Figure 4 highlights the portions of the *E. coli* pol I tertiary structure corresponding to the three diagnostic motifs. Motif A, characterized by a conserved Asp and Ser, corresponds to strand D and helix L. Motif B, with conserved Lys, Tyr and Gly encompasses most of helix O and the following long loop region. Motif C delineates strands I2 and I3 with the mostly conserved negative charges contained in the connecting loop of the β-hairpin. Since the N-terminal Asp of this doublet is the only one universally conserved, it is presumed to be catalytically most important. It is clear that the structural segments A and B are likely to come into contact with DNA and that the strand I2–I3 hairpin loop could place at least one catalytic residue in the polymerase active site. Model building of DNA into the Klenow fragment structure also supports the importance of these regions (Ollis *et al.*, 1985a,b; Warwicker *et al.*, 1988). Furthermore, the C_α atom of motif C Asp882 is within 3.5 Å of that from Asp705, conserved in motif A (C_α atom coordinates given in Brookhaven database (Bernstein *et al.*, 1977, file 1DPE). Their spatial proximity would allow both to participate in catalysis. Region 2 of pol β (sequence TCR between strands I and L – see Figure 2 and 4) is also located in the vicinity of this area but no sequence homology with the pol α type could be found for this region. Region 1 fails in an undefined part of the electron density map (see Ollis *et al.*, 1985a, Figure 4).

From structure to function. Corroborative biochemical evidence points to the importance of these three motifs in the DNA polymerase activity. A synthesized *E. coli* pol I oligopeptide corresponding to the N-terminal most two-thirds of the loop region connecting helices I3 and F (motif B – see Figure 4) has been shown to bind strongly to the triphosphate substrate of pol I as well as to single- DNA (Modak, 1986). Furthermore, photo-affinity labelling with 8-azido-dATP identifies Tyr766 as a residue in the active site (Joyce and Steitz, 1987) while Lys758, also part of this motif, succumbs to chemical labeling. Site-directed mutagenesis at phosphate (Basu and Modak, 1987). A further synthesized peptide (Bhattacharyya *et al.*, 1988; Modak, 1988) corresponding to helix G and strands I2 and I3 (motif C) was not bound by itself the pol I substrate, although it spontaneously entered its proper folding; however, this motif does not exhibit that this peptide, although unable to bind dNTP on its own, can co-operate with other regions of the native protein and be just part of the binding site. In fact, mutant of pol I, which is in the loop joining strands I2 and I3 and sequentially adjacent to the conserved Asp882 of motif C, has been shown to be involved in the binding of [^{32}P]TTP (Pandey *et al.*, 1987). In addition, the beginning of helix Q, which is close in space to the end of helix Q (motif B) and results A and C, can also be labeled by photoaffinity (Modak *et al.*, 1988).

Similarly, for the pol α-type, genetic studies of the Herpes simplex virus polymerase (Larder *et al.*, 1987a) revealed that four out of the six identified drug-resistant polymerase mutants cluster in motifs A and B. The drug used was a dNTP analog. Other mutants associated to drug and substrate recognition were also mapped by Gibbs *et al.* (1988) in motifs A and B. This is in agreement with a mutant role of these motifs in dNTP binding for the pol α type, as in the case of the pol I type.

Finally, for no-gap DNA polymerases type (pol β), a chemical affinity labeled [^{32}P]8-azido-dATP peptide on the terminal transferase allowed the mapping of the dNTP binding site; this peptide contained the sequence GHDVD that is part of motif C (Evans *et al.*, 1986) and that resembles both YGIITD and VSIDE. This supports the alignment given in Figure 3, even though motif B could not be found with certainty in this pol β type of DNA polymerases. Furthermore, the number of residues between motifs A and C in the terminal transferase (~ 150) matches well the corresponding number in the Klenow pol I, averaging about 130 residues.

Structural implications of the distance variability between motifs. Figure 3 shows the number of residues contained between each of the motifs for all the DNA pol α sequences and for the Klenow polymerase I domains. Between motifs A and B a comparable number of residues is found for both pol I and α types, while between B and C the pol α sequences generally contain considerably fewer amino acids than their putative Klenow counterparts. This region of the structure encompasses helix F, strands I2 and I3, and helix Q. Apart from the beginning of helix Q, these regions are unlikely to be in contact with the DNA, according to model building studies (Warwicker *et al.*, 1988; Figure 4), suggesting that this region could be consistently

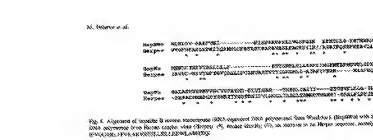


Fig. 4. Alignment of motifs B across nucleotide (RNA-dependent RNA polymerase) from Woodchuck (Marmota monax) hepatitis virus (WHV) and DNA polymerases from Hepatitis B virus (HBV) and Cauliflower mosaic virus (CaMV). Conserved residues are shown in bold. Secondary structure elements are indicated above the sequences. Numbering of residues is given at the end of each line. Sequences are from the following sources: WHV, WHVSURG; HBV, HBVADW2; CaMV, CAMVDNAVI.

cations) without apparent catalytic harm. The shortest segment between motifs B and C is 29 amino acids in the human pol α sequences; a peptide of this size could easily connect Ile779 at the C terminus of the Klenow motif B to Arg815 at the N-terminus of motif C, i.e. connect the loop following beta O with strands 12 and 13 of motif C. In fact, given the 4Å distance between the C_α atoms of Ile779 and Arg815 in the T7 pol I tertiary structure (Ollis et al., 1985b), 11 residues in helical conformation (strain C_α distance of 1.5 Å) and eight in a coil structure (strain C_α distance of 2.9 Å) could span the segment of 4 Å. This helical model would favor the maintenance of the N-terminal region of motif C, whose sequence is also reasonably conserved in the six type 1 DNA polymerases sequenced (see Figure 2).

The distance between motifs A and B can also vary from 30 to 120 residues in the pol α family; this region corresponds to helices M and N of the Klenow structure. In the pol I family, this distance ranges from 20 to 40 residues. An inspection of Figure 4 suggests the possibility of insertions in this region.

DNA-dependent RNA polymerases

In the family, three conserved regions that matched the three motifs A, B and C could be found. Insertion residues are maintained and the residue lengths between them compare well with those of DNA-dependent DNA polymerases. However, since this family is composed of four members with very two sufficiently distant ones, more sequences are needed in order to narrow the number of conserved regions that can truly functionally connected.

RNA-dependent polymerases

Of the four motifs of viral RNA-dependent RNA polymerases, as well as reverse transcriptase sequences recently identified by Poch et al. (1989), two match motifs A and C of DNA-dependent polymerases. These motifs are the only sequence features shared by all these sequences. The 31% identity alignment of the hepatitis B reverse transcriptase from Woodchuck (Gerlich et al., 1982) with the Rauscher murine virus DNA pol α (Fabris et al., 1983) strongly supports the relationships between DNA- and RNA-dependent polymerases. The closeness of these two sequences is also sensible on evolutionary grounds, because reverse transcriptases have recently been postulated to be an essential late ancillary step in going from an RNA-dependent RNA polymerase to a DNA-dependent DNA polymerase (Lazcano et al., 1988). Furthermore, it should be pointed out the hepatitis B viruses are the only viruses encoding reverse-transcriptase activity that have a DNA genome.

A site-directed mutagenesis experiment has pointed to the catalytic importance of the Asp—Asp doublet in motif C of RNA-dependent polymerases (Inokuchi and Hirashima, 1987). In addition, site-specific mutagenesis of the hepatitis mutagenesis deficiency virus reverse transcriptase revealed that mutation of the first conserved Asp of motif C into Asn completely destroyed activity (Larder et al., 1987b). The mutation of the other strictly conserved Asp residue in motif A into Glu had a similar effect.

The sequence length between motifs A and C in the RNA-dependent polymerases averages ~ 30 residues, which is considerably shorter than the 150–170 amino acids in pol I, though not sequence from influenza viral cousins 718 amino residues, comparable with some pol αs (Figure 3). Once again helix P and part of the helix Q and strands 10 and 11 are possible candidates for deletions. The absence of all the distances O8 residues suggests that structure can survive in the M and N helices and in the loop following helix O, as they family.

Perspectives and conclusions

Argos (1988) previously discussed the conservation of motif C in pol αs and RNA-dependent polymerases; we extend this observation to the pol βs, pol γs and some DNA-dependent RNA polymerases. Apparently the fully conserved first Asp of the Asp—Asp doublet of motif C is evolutionarily more important than the second. We also located that there is at least another motif upstream (motif A), with another strictly conserved aspartate located in a turn between a beta strand and an alpha helix. In spectral proximity of motif C, genetic experiments suggest that this motif might also be involved in the catalytic process. One of the many possible functional roles for these residues is that the strictly conserved Asp residue in motif C may co-operate with the one of motif A to bind a magnesium ion that would be part of the dNTP binding site. This site would be located at the bottom of a cleft containing the DNA. This cleft, similar to the one of the Klenow fragment, is already apparent in a 4 Å map of T7 DNA-dependent RNA polymerase (Jeon et al., 1988). If this is true, the diffences in rates of polymerases could be due to variations of divergence from a common structure, with essentially the same folding as the counterparts of conserved residues in the same spatial arrangement in the complex site could be due to convergent evolution, as observed in nine nucleotide and ribonuclease active site.

In spite of wide apparent sequence variability (likely to reflect a very ancient divergence), it is very tempting to analyze the strong, structurally unifying, protein sequence data many polymerases may hold like the known tertiary architecture of the E.coli pol I. Experiments with this hypothesis in the experimental observation that it is possible to change the template or substrate specificity of certain polymerases (type I Mg^{2+} is replaced by Mn^{2+} (see Lazcano et al., 1988). This model, however, should be viewed as speculative, even though several lines of evidence point in this unifying conclusion. We believe it deserves attention, because site-directed mutagenesis experiments aimed at probing the catalytic site and template specificity should benefit from our hypothesis, which gives a possible structural framework for

factors: supersequences. If repeated episodic or a heterozygotes of the reverse transcriptase but human immunodeficiency virus group by be differences months (Lewis et al., 1990b) and if the high resolution structure of T7 RNA polymerase soon become available, therefore first rule for the hypothesis made been should be forthcoming.

References

Argos,P. (1985) *EMBO J.* 4, 1273 - 1280

Argos,P. (1987) *J. mol. Biol.*, 197, 331 - 348

Argos,P. (1988) *Nucl. Acids Res.*, 16, 9909 - 9916

Argos,P., Tsernoi,V.S. and Hillenjan,J. (1986) *Progress* 169, 261 - 276.

Bass,A. and Weintraub,H. (1987) *Biochemistry*, 26, 1780 - 1790.

Bossad,A., Dimitris,A., Julia,M. and Nunes,S. (1987) *EMBO* 6, 4019 - 4025.

Bernad,A., Blanco, L., Lazaro J.M. / Martin,G. and Salas,M. (1989) Cell, 59, 219 - 228.

Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.E., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.* 112, 535 - 542

Bonnet,F.X., Bruch,C.W. and Cramton,M.E. (1989) *Nucleosides*, 28, 721 - 730.

Bonner,T., Gilo,D.A., Sosa,T.A., *and Bonner,W.* (1986) *Proteins*, 1, 69 - 75.

Bothner,B., Hsu,C.H. and Stewart,R. (1987) *J. Virol.*, 61, 31 - 35.

Bolle,H.L., Unge,T.C., Pohl,E.R., Hurstad,O.W., Rasmussen,M.I., Wilde,K.W. and Cox,J.M. (1985) *Proc. Natl. Acad. Sci. USA*, 82, 2969 - 2973.

Bolle,J., Chou,R., Sasson,A., Cheng,T.C. and Cheng,Y.C. (1989) *Proc. Natl. Acad. Sci. USA*, 86, 6972 - 6976.

Chisholm,W. and Burgess,R.R. (1990) *Nucl. Acids. Res.*, 18, 6745 - 6753

Coulter,N.T. and Hendrickson,A. (1989) *J. Virol.*, 63, 3960 - 3969.

Daniel,M. and Weiss,P.A. (1987) *Trends Biochem. Sci.*, 12, 273 - 275

Deng,H., Lacroix,M.C., Bobek,J.C. and Bell,J.C. (1990) *Proc. Natl. Acad. Sci. USA*, 87, 6287 - 6291

Kamer,G., *and Argos,P.* (1984) *Nucl. Acids Res.*, 12, 7269 - 7282.

Kerwalla,S., Nakata,T., Hagiwara,T., Kuroda,Y., Yoshida,S. *and Noda,K.* (1989) *Nucl. Acids Res.*, 16, 5777 - 5790

Lander,M.A., Sabo,L.B.D. and Sa-Nogueira,I. (1987) *EMBO J.*, 6, 349 - 357.

Lazaro,J.M., Bernad,J.M., Blanco,L., Salas,M., and Bonfy,S. (1989) *Nucleic Acids Res.*, 17, 9709 - 9717

Lazcotti,P.C., Smith,R., Smith,K.K., Shamblin,K., Greenwood,R. and Chilcote,D.D. (1989) *J. Biol. Chem.*, 264, 16477 - 16477

Lazcano,A., Fox,G., Oro,J.*, Barghoorn,* Bonner,G. and Oro,J. (1988) *J. Mol. Evol.*, 27, 365 - 376

Lazcano,A., Oro,J. (1985) *Proc. Natl. Acad. Sci. USA*, 82, 4942 - 4945.

Lazcano,A., Guerrero,R., Margulis,L., Oro,J. (1988) *J. Mol. Evol.*, 27, 365 - 376

Lazcano,A., Miller,S.L., Orenberg,J., Deamer,D.W. (1988) *Biochemistry*, 27, 4987 - 4993

Lazcano,A., Watanabe,C., Bailey,S.K., Lanthier,K.A., Powell,K.L., Nambu,J.M., Farber,M. and Bowman,D.R. (1986) *Biochemistry*, 27, 4987 - 4993

Mahadevan,S., Hamasaki,K., Liu,T., Ueki,T. and Yamagami,M. (1990) *J. Biol. Chem.*, 265, 6960 - 6962.

Mattes,R.S., Bishop,E.L. and Dayhoe,D.A. (1988) *Cell*, 55, 83 - 90.

Mikawa,S. (1989) *FASEB J.*, 3, 1703 - 1710.

Mohan,P., Rose,A., Sharman,T.J. and Mirfakhrai,M.J. (1988) *Biochemistry*, 27, 276 - 278.

Ohta,S., Bock,Z., Kim,Jin-S., Wong,N.G. and Seibe,T.A. (1990) *Nature*, 345, 765 - 768.

Oshy,D.A., Kline,D. and Sabo,T.A. (1989) *Nature*, 337, 787 - 795.

Reddy,V.N., Wilkinson,S., Rose,K.L. and Nadas,M.L. (1987) *Biochemistry*, 26, 7746 - 7748.

Peterson,R.D., Cheng,L.C., Wickliffe,J.K., Niles,G.J., Cheng,L.M.S. and Bellamy,F.J. (1989) *J. Biol. Chem.*, 264, 10950 - 10962.

Poch,O., Kawaguchi,I., Ortoue,M. and Seeburgh,(1989) *J.*, 8, 3867 - 3874

Poch,O. and Bellamy,J. (1986) *J. Mol. Ed.*, 7 - 15.

Rasmussen,Jon,P., Mohr,G.P. and Morrow,C.D. (1988) *FASEB J.*, 2, A305.

Rose,J.K., Rose,J.P., Cheng,Y.T., Lah,J.M. and Hsu,C.H. (1988) *Proteins*, 3, 256 - 266. and *Proceedings of the Annual Meeting of the American Crystallographic Association* 3, 1988, (1988), p.116, p.116 (1987).

Young,L.S.F., Young,D.W., and Kenedy, (1989) *FASEB J.*, 3, 19 - 25.

Zimmern,D., Kaesberg,P., Hollingsworth,D., Dolnick,F., Kumar,P. and Wilson,I.H. (1986) *Proc. Natl. Acad. Sci. USA*, 83, 2486 - 2490.

Received on November 30, 1990, accepted on February 24, 1990

EXHIBIT B

Docket No.: NEB-166-PUS

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANTS: Jack et al. EXAMINER: Hutson
SERIAL NO.: 10/089,027 ART UNIT: 1652
DATE FILED: March 26, 2002
TITLE: Incorporation of Modified Nucleotides By Archaeon DNA Polymerases And Related Methods

Mail Stop AF
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION UNDER 37 C.F.R. §1.131

As a below named inventor, I hereby declare that:

1. My name is Dr. William Jack, Research Director for the DNA Enzymes Division at New England Biolabs Inc. My resume is attached.

2. I have been studying the structure and function of DNA polymerases for over 16 years.

3. I was a member of the group of scientists at New England Biolabs that isolated, characterized, and cloned the first hyperthermophilic archaeal DNA polymerase. Our continuing work with archaeon DNA polymerases identified a surprisingly homogeneous set of enzymes. We claimed this group of DNA polymerases in US Patent 5,500,363. In this patent, the United States Patent and Trademark Office recognized the validity of our claim to a class of archaeon DNA polymerases defined by the DNA encoding the enzyme and its

ability to hybridize under defined conditions to various specified DNA sequences. The group was exemplified by T.litoralis (Vent), GBD (Deep Vent), and 9°N DNA Polymerases.

4. We also found that this group of polymerases had a high degree of amino acid sequence identity. A comparative three-dimensional alignment of members of this group of enzymes showed a high degree of structural conservation, consistent with the observed high degree of primary amino acid sequence identity/similarity. See for example, Vent (Rodriguez, et al., 2000), Tgo (Hopfner, et al., 1999), D. Tok (Zhao, et al., 1999), and KOD (Hashimoto, et al., 2001) DNA Polymerases.

5. The structural equivalence of this group of polymerases is further supported by experiments reported in Example 10 of the above application in which we show that mutation of an analogous residue in Vent and 9°N DNA Polymerases yields enzymes with equivalent acyclonucleotide incorporation efficiencies.

6. We discovered that this group of enzymes is capable of efficiently utilizing acyclonucleotides as substrates. We demonstrated this property using four examples of polymerases within this tightly defined group. Any molecular biologist of ordinary skill in the art would expect from these findings that this property would occur in all members of the enzyme group defined above.

7. Additionally, my colleagues and I have published articles in peer reviewed journals discussing the physical basis for the preferential incorporation of acyclonucleotides, and also for the enhanced incorporation with Vent A488L and 9°N A485L DNA Polymerase mutants. See Gardner, et al. (2004) on page 11841, column 1, paragraph 2 and page 11841, column 2, paragraph

1, respectively.

8. I assert that the combination of the high degree of homogeneity in DNA and amino acid sequences of archaeon DNA polymerases, plus the structural evidence that modification of specific amino acids alters enzyme specificity, would be sufficient to assure a person of ordinary skill in the art that the class of polymerases as defined above will interact with acyclonucleotide substrates as shown in the above application.

9. To further support the above statements, we have conducted additional experiments to confirm that archeon Family B polymerases with an amino acid sequence identity of greater than 30% can utilize acyclonucleotides as a substrate. This data is attached to the present declaration as appendix 1.

9. I further declare under penalty of perjury pursuant to laws of the United States of America that the foregoing is true and correct and that the Declaration was executed by me on:

William E. Jack

Date: 4 May 2006

References:

A. Gardner, C.M. Joyce, W.E. Jack (2004) J. Biol. Chem. 279, 11834-11842.

H. Hashimoto, M. Nishioka, S. Fujiwara, M. Takagi, T. Imanaka, T. Inoue, Y. Kai (2001) J. Mol. Biol. 306, 469-477.

K.P. Hopfner, A. Eichinger, R.A. Engh, F. Laue, W. Ankenbauer, R. Huber, B. Angerer (1999) Proc. Natl. Acad. Sci. USA 96, 3600-3605.

A.C. Rodriguez, H.W. Park, D. Mao, L.S. Beese (2000) J. Mol. Biol. 299, 447-462.

Y. Zhao, D. Jeruzalmi, I. Moarefi, L. Leighton, R. Lasken, J. Kuriyan (1999) Structure 7, 1189-1199.

William Erle Jack
New England Biolabs
240 County Road
Ipswich, MA 01938
(978) 380-7257
(978) 921-1350 (fax)
email: jack@neb.com

RESEARCH INTERESTS

Enzymatic and structural aspects of protein-nucleic acid interactions. Thermostable DNA polymerase kinetics and function.

RESEARCH EXPERIENCE

New England Biolabs (Beverly, MA).

2005-present — Division Head, DNA Enzymes

1987-present — Senior Staff Scientist

Research: Kinetic characterization of thermostable DNA polymerases. Creation and characterization of DNA polymerase variants with altered substrate recognition. Over-expression and characterization of restriction and modification enzymes.

2000-present — New England Biolabs Institutional Biosafety Committee Chair

Rockefeller University (NY, NY) Laboratory of Biochemistry and Molecular Biology.

1983-1987 — Postdoctoral Fellow in the laboratory of R.G. Roeder.

Research: Structural and functional characterization of wild type and mutant forms of *Xenopus* RNA polymerase III transcription factor A. Glucocorticoid hormone-induced transcription enhancement *in vitro*.

Duke University (Durham, NC) Department of Biochemistry.

1977-1983 — Graduate Student in the laboratory of P. Modrich.

Research: Kinetics and thermodynamics of DNA site location, recognition and cleavage by *EcoRI* endonuclease.

EDUCATION

Doctor of Philosophy (Biochemistry), Duke University, 1983 (Paul Modrich, advisor).
Bachelor of Arts (Chemistry), *Magna Cum Laude*, University of Utah, 1977.

TRAINING

2000 — Sixth National Symposium on Biosafety: Prudent Practices for the New Millennium (Conducted by the Centers for Disease Control and Prevention)

PUBLICATIONS

"Comparative Kinetics of Nucleotide Analog Incorporation by Vent DNA Polymerase," Andrew F. Gardner and **William E. Jack**, *J. Biol. Chem.* 279, 11834-11842 (2004).

"Acyclic and Dideoxy Terminator Preferences Denote Divergent Sugar Recognition by Archaeon and Taq DNA Polymerases," Andrew F. Gardner and **William E. Jack**, *Nucleic Acids Res.*, 30, 605-613 (2002).

"The Kinetic Mechanism of *Eco*RI Endonuclease," David J. Wright, **William E. Jack** and Paul Modrich, *J. Biol. Chem.* 274, 31896-31902 (1999).

"Determinants of Nucleotide Sugar Recognition in an Archaeon DNA Polymerase," Andrew F. Gardner and William E. Jack, *Nucleic Acids Res.* 27, 2545-2553 (1999)

"Photochemically Initiated Protein Splicing", Sandra N. Cook, **William E. Jack**, Xiaofeng Xiong, Lora E. Danley, Jonathan A. Ellman, Peter G. Shultz, and Christopher J. Noren, *Angew. Chem. Int. Ed. Engl.* **34**, 1629-1630 (1995).

"Three-dimensional Structure of the Adenine-specific DNA Methyltransferase M.TaqI in Complex with the Cofactor S-adenosylmethionine", J. Labahn, J. Granzin, G. Schluckebier, D.P. Robinson, **W.E. Jack**, I. Schildkraut, and W. Saenger, *Proc. Natl. Acad. Sci. USA* **91**, 10957-10961 (1994).

"Expression, Purification, and Crystallization of Restriction Endonuclease PvuII With DNA Containing Its Recognition Site", K. Balendiran, Joseph Bonventre, Roger Knott, **William Jack**, Jack Benner, Ira Schildkraut, and John E. Anderson, *Proteins: Structure, Function, and Genetics* **19**, 77-79 (1994)

"Protein Splicing Elements: Inteins and Exteins-A Definition of Terms and Recommended Nomenclature", Francine B. Perler, Elaine O. Davis, Gary E. Dean, Frederick S. Gimble, **William E. Jack**, Norena Neff, Christopher J. Noren, Jeremy Thorner and Marlene Belfort, *Nucleic Acids Res.* **22**, 1125-1127 (1994).

"Characterization of a DNA Polymerase from the Hyperthermophile Archaea *Thermococcus litoralis*", Huimin Kong, Rebecca B. Kucera and **William E. Jack**, *J. Biol. Chem.*, **268**, 1965-1975 (1993)

"Protein Splicing Removes Intervening Sequences in an Archaea DNA Polymerase", Robert A. Hodges, Francine B. Perler, Christopher J. Noren and **William E. Jack**, *Nucleic Acids Res.* **20**, 6153-6157 (1992).

"Intervening Sequences in an Archaea DNA Polymerase Gene", Francine B. Perler, Donald G. Comb, **William E. Jack**, Laurie S. Moran, Boqin Qiang, Rebecca B. Kucera, Jack Benner, Burton E. Slatko, Donald O. Nwankwo, S. Kay Hempstead, Clotilde K.S. Carlow and Holger Jannasch, *Proc. Natl. Acad. Sci. USA* **89**, 5577-5581 (1992).

"Overexpression, Purification and Crystallization of *Bam*HI Endonuclease", **William E. Jack**, Lucia Greenough, Lydia F. Dorner, Sloung-yong Xu, Teresa Szreiecka, Aneel K. Aggarwal and Ira Schildkraut, *Nucl. Acids Res.* **19**, 1825-1829 (1991).

"Nucleotide Sequence of the *FokI* Restriction-Modification System: Separate Strand-Specificity Domains in the Methyltransferase", Mary C. Looney, Laurie S. Morse, **William E. Jack**, George R. Feehery, Jack S. Benner, Barton E. Slatko and Geoffrey G. Wilson, Gene **80**, 193-208 (1989).

"M.FokI methylates adenine in both strands of its asymmetric recognition sequence", David Landry, Mary C. Looney, George R. Feehery, Barton E. Slatko, **William E. Jack**, Ira Schildkraut, and Geoffrey G. Wilson, *Gene* **77**, 1-10.

"Mechanism of specific site location and DNA cleavage by *EcoRI* endonuclease", Brian J Terry, **William E. Jack** and Paul Modrich, *Gene Amplif Anal.* volume 5, pp. 105-118 (1987)

"Facilitated diffusion during catalysis by EcoRI endonuclease. Nonspecific interactions in EcoRI catalysis", Brian J. Terry, **William E. Jack** and Paul Modrich, *J Biol. Chem.* **260**: 13130-13137 (1985).

Participation of Outside DNA sequences in the EcoRI Endonuclease Reaction Pathway, **William E. Jack**, dissertation (Duke University, Durham, NC) (1983).

"Thermodynamic parameters governing interaction of EcoRI endonuclease with specific and nonspecific DNA sequences", Brian J. Terry, **William E. Jack**, Robert A. Rubin and Paul Modrich, *J Biol. Chem.* , **258**, 9820-9825 (1983).

"Involvement of outside DNA sequences in the major kinetic path by which EcoRI endonuclease locates and leaves its recognition sequence", **William E. Jack**, Brian J Terry, and Paul Modrich, *Proc Natl. Acad Sci U S A* **79**: 4010-4014 (1982)

"DNA determinants important in sequence recognition by *EcoRI* endonuclease", A-Lien Lu, **William E. Jack** and Paul Modrich, *J Biol. Chem.* **256**: 13200-13206 (1981).

"Structures and Mechanisms of *EcoRI* DNA Restriction and Modification Enzymes", **William E. Jack**, Robert A. Rubin, Andrea Newman and Paul Modrich, *Gene Amplif. Anal.* volume 1, p 165-179 (1981).

US PATENTS

"Use of rare-specific nicking endonucleases to create single-stranded regions and applications thereof," **William E. Jack**, Ira Schildkraut, Julie Forney Mertes, US Patent 6,660,475, December 9, 2003.

"Recombinant thermostable DNA polymerase from archaebacteria," Donald G. Comb, Francine Perler, Rebecca Kucera, **William E. Jack**, US Patent 5,834,285, November 10, 1998.

"Modified protein comprising controllable intervening protein sequences or their elements methods of producing same and methods for purification of a target protein comprised by a modified protein," Donald G. Comb, Francine B. Perler, **William E. Jack**, Ming-Qun Xu, Robert A. Hodges, Christopher J. Noren, Shaorong S.C. Cheng, Eric Adam, Maurice Southworth, US Patent 5,834,247, November 10, 1998.

"Recombinant thermostable DNA polymerase from archaebacteria," Donald G. Comb, Francine Perler, Rebecca Kucera, **William E. Jack**, US Patent 5,500,363, March 19, 1996.

"Modification of protein by use of a controllable intervening protein sequence," Donald G. Comb, Francine B. Perler, **William E. Jack**, Ming-Qun Xu, Robert A. Hodges, US Patent 5,496,714, March 5, 1996.

"Recombinant thermostable DNA polymerase from archaebacteria," Donald G. Comb, Francine Perler, Rebecca Kucera, **William E. Jack**, US Patent 5,352,778, October 4, 1994.

"Purified thermostable DNA polymerase obtainable from thermococcus litoralis," Donald G. Comb, Francine Perler, Rebecca Kucera, **William E. Jack**, US Patent 5,322,785, June 21, 1994

doi:10.1016/j.jmb.2009.07.028 Available online at http://www.sciencedirect.com on [illegible] J. Mol. Biol. (2009) 393, 447–462

JMB

Crystal Structure of a Pol α Family DNA Polymerase from the Hyperthermophilic Archaeon *Thermococcus sp.* 9°N-7

A. Chapin Rodriguez†, Hee-Won Park†, Chen Mao and Lorena S. Beese*

Department of Biochemistry, Duke University Medical Center, Durham, NC 27710, USA

The 2.25 Å resolution crystal structure of a pol α family (family B) DNA polymerase from the hyperthermophilic marine archaeon *Thermococcus* sp. 9°N-7 (9°N-7 pol) provides new insight into the mechanism of pol α family polymerases that include essentially all of the eukaryotic replicative and viral DNA polymerases. The structure is folded into NH_2-terminal, editing 3′–5′ exonuclease, and polymerase domains that are topologically similar to the two other known pol α family structures, bacteriophage RB69 and the recently determined *Thermococcus* gorgonarius, but differs in their relative orientations and conformations.

The 9°N-7 polymerase domain structure is reminiscent of the "closed" conformation characteristic of ternary complexes of the pol I polymerase family obtained in the presence of free dNTP and DNA substrates. In the apo-9°N-7 structure, this conformation appears to be stabilized by an ion pair. Thus far, the other apo-pol α structures that have been determined adopt open conformations. These results therefore suggest that the pol α polymerases undergo a series of conformational transitions during the catalytic cycle similar to those proposed for the pol I family. Furthermore, comparisons of the orientations of the fingers and exonuclease (exo)domains relative to the palm subdomain that contains the pol active site suggests that the exonuclease domain and the fingers subdomain of the polymerase can move as a unit and may do so as part of the catalytic cycle. This provides a possible structural explanation for the interdependence of polymerization and editing exonuclease activities unique to pol α family polymerases.

We suggest that the DP-terminal domain of 9°N-7 pol may be structurally related to an RNA-binding motif, which appears to be conserved among archaeal polymerases. The presence of such a putative RNA-binding domain suggests a mechanism for the observed misincorporation of bacteriophage T4 DNA polymerase synthesis by binding to its own mRNA. Furthermore, conservation of this domain could indicate that such regulation of pol expression may be a characteristic of archaea. Comparison of the 9°N-7 pol structure to its unstable homolog from bacteriophage RB69 suggests that thermostability is achieved by shortening loops, forming two disulfide bridges, and increasing electrostatic interactions at subdomain interfaces.

Keywords: Archaea; X-ray structure; replication; exonuclease; family B DNA polymerase

*Corresponding author

†Contributed equally to the manuscript.

Abbreviations used: pol, polymerase; T_m, *Thermococcus gorgonarius*; ddNTP, dideoxyribonucleotide.

E-mail address of the corresponding author: sbeese@neuro.duke.edu

0022-2836/$ - see front matter

Introduction

DNA polymerases catalyze the template-directed addition of nucleotides onto the 3′-OH group of the DNA primer terminus. These enzymes replicate DNA with the required accuracy essential for genomic

mic stability, but generate sufficient mutations to stimulate and maintain evolution. Unlike Eucarya and Bacteria, relatively little is known about DNA replication in *Archaea* (Perler *et al.*, 1992), one of the three major evolutionary lineages of life (Woese *et al.*, 1990). Archaea play a significant role in the biosphere, accounting for up to 30% of the biomass in certain Antarctic waters (De Long *et al.*, 1994), and exhibit much greater diversity than had originally been suspected (Burns *et al.*, 1996). Many characterized archaeal species are adapted to live in environments of extreme temperature, pressure, salinity, and/or pH such as hydrothermal vents, and hot springs (Ross & Adams, 1995).

Although archaeal cells share many morphological features with Eucarya, archaeal proteins involved in gene expression including DNA replication, transcription, and translation have been found to be similar to those from Eucarya (Edgell & Doolittle, 1997; Puhl *et al.*, 1996) in particular most of the archaeal DNA polymerases that have been sequenced belong to the α-like polymerase family (family B) that includes essentially all the eukaryotic replicative and viral DNA pols (Braithwaite & Ito, 1993; Edgell *et al.*, 1997).

Crystal structures exist for DNA pols from each of four families pol I (family A), pol α (family B), pol β (family X) and reverse transcriptase (retroviral by type & Steitz, 1994; Doublié *et al.*, 1998). Although pols from different families are structurally quite diverse, several common features have emerged. The pol domains from each resemble a right hand and may be further divided into palm, fingers, and thumb subdomains, as was originally described for the large fragment of *Escherichia coli* pol I (Klenow fragment (Ollis *et al.*, 1985). All polymerases appear to share the same mechanism for nucleotidyl transfer involving two divalent metal ions (reviewed by Beese & Steitz, 1994). In addition, based on structures containing DNA and dNTP bound to pols from pol I, pol β, and reverse transcriptase inhibitors, a conformational change in the fingers subdomain from an open to a closed conformation is proposed to occur during the catalytic cycle (reviewed by Doublié *et al.*, 1999).

The pol α family polymerases are of medical importance as targets for development of antiviral and antitumor therapeutics. For example, human pol α is a target in the treatment of acute myelogenous leukemia and chronic lymphocytic leukemia (Keating *et al.*, 1982; Robertson & Plunkett, 1985) and a variety of nucleotide analogs with antitumor activity inhibit strand elongation by pol α (Huang & Plunkett, 1991; Cazaubon & Plunkett, 1995). Furthermore, polymerases, particularly those that are thermostable, have a number of critical biotechnological applications ranging from PCR to cloning and DNA sequencing. Despite their biological, medical and biotechnological importance, the pol α class of polymerases has not been structurally as well characterized as other DNA polymerase families.

Here we report the 2.25 Å resolution crystal structure of a pol α family DNA polymerase from the hyperthermophile marine archaeon *Thermococcus* sp. 9°N-7 (9°N-7 pol). *Thermococcus* sp. 9°N-7 was isolated from a 9-gdegree-north vent at 9° N latitude off the East Pacific Rise (Southworth *et al.*, 1996). The structure is folded into NH$_2$-terminal, editing 3′-5′ exonuclease, and polymerase domains that are topologically similar to the two other known pol α family structures Bacteriophage RB69 (Wang *et al.*, 1997) and the recently determined *Thermococcus gorgonarius* (Tgo) (Hopfner *et al.*, 1999), but differs in their relative orientation and conformation.

The pol domain structure is reminiscent of the "closed" conformation characteristic of ternary complexes of the pol I polymerase family obtained in the presence of their dNTP and DNA substrates. In the apo-9°N-7 structure, this conformation appears to be stabilized by an ion pair. Thus far, the two other apo-pol α structures that have been determined adopt open conformations. These results therefore suggest that the pol α polymerases undergo a series of conformational transitions during the catalytic cycle similar to those proposed for the pol I family. Furthermore, comparison of the architectures of the fingers and exonuclease domains relative to the palm subdomain that contains the pol active site suggests that the exonuclease domain and the fingers subdomain of the polymerase can serve as a unit, and may do so as part of the catalytic cycle. This provides a possible structural explanation for the interdependence of polymerization and editing exonuclease activities unique to pol α family polymerases.

We suggest that the NH$_2$-terminal domain of 9°N-7 pol is structurally homologous to the RadA/Sms RNA-binding motif with an exposed patch of aromatic amino acid residues that recognize the DNA pol, which is homologous to 9°N-7 pol, is known to bind its own mRNA and repress its own synthesis. The homology relationships in the RNA-binding motif suggest a structural basis for this regulatory mechanism. Furthermore, the conservation of this domain in other archaeal pols suggests that such autogenous regulation of pol expression may be general for archaea.

Results and Discussion

Crystal structure of *Thermococcus* sp. 9°N-7 pol

The structure of the full-length, 775-residue enzyme showing the double mutation D141A and D143A(s) was determined using the multiple isomorphous replacement method to a resolution of 2.25 Å. The current model has an R factor of 23.9% (R_{free} = 26.8 %) (Table 2). A Ramachandran plot of the model shows 86.5% of the residues in the most favored region and the remainder in additional allowed regions (12.4%) and generously allowed regions (0.8%). A total of 37 residues are not traced in the model and lie in regions of poorly defined electron density. The first of these gaps

Table 1. Crystallographic data collection and refinement statistics

Data	Native[a]	Native	Native[b]	TMA+	TMA+	TMA2+	TMA2+	TMA3+	TMA3+	TMA3+	TMA3+
A. Data collection and phasing[c]											
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P4_12_12$
a, c (Å)	89.1, 207.3	88.9, 206.7	89.1, 207.3	89.1, 207.1	89.1, 207.1	89.1, 207.1	89.1, 207.1	89.0, 206.9	89.0, 206.9	89.0, 206.9	89.0, 206.9
Resolution (Å)	47.5–2.3	50–2.5	50–2.3	50–2.6	50–2.6	50–2.5	50–2.5	50–2.5	50–2.5	50–2.5	50–2.5
Measured reflections	1,054,856	1,134,610	1,139,554	1,134,610	1,134,610	1,134,610	1,134,610	1,134,610	1,134,610	1,134,610	1,134,610
Unique reflections	89,695	82,488	89,328	69,295	69,295	69,295	69,295	69,295	69,295	69,295	69,295
Completeness (%)	99.9 (99.9)	99.9 (99.9)	99.9 (99.9)	99.9 (99.9)	99.9 (99.9)	99.9 (99.9)	99.9 (99.9)	99.9 (99.9)	99.9 (99.9)	99.9 (99.9)	99.9 (99.9)
R_{sym} or R_{merge}	0.074	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
$I/\sigma I$	17.7	11	11	11	11	11	11	11	11	11	11
Phasing power[d]											
B. Refinement (1.9 Å)[e]											
Resolution (Å)	47.5–1.9										
Reflections (work/free)	85,125/4,570										
R_{cryst}/R_{free} (%)	19.2/23.3										
No. atoms	8,140										
Protein	7,128										
Ligand/ion	28										
Water	984										
B-factors (Å²)	23.9										
Protein	22.9										
Ligand/ion	31.8										
Water	34.9										
R.m.s. deviations											
Bond lengths (Å)	0.006										
Bond angles (°)	1.1										

[a] Data collected at 100 K. [b] Data collected at room temperature. [c] Values in parentheses are for highest-resolution shell. [d] Phasing power = $\langle F_H \rangle / \langle E \rangle$, where $\langle F_H \rangle$ is the average structure factor amplitude of the heavy-atom reflections and $\langle E \rangle$ is the residual lack of closure. [e] $R_{cryst} = \sum ||F_o| - |F_c|| / \sum |F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively. R_{free} is the same as R_{cryst} but for 5% of the reflections chosen at random and omitted from refinement.

cerase at the bottom of the palm domain (residues 568–575), and the remainder are within the thumb regions that are frequently observed to be partially disordered in apo polymerase structures, as is also the case here (e.g. Ollis et al., 1995; Kiefer et al., 1997). Although no disulfide bridges were included in the refinement, four C_β residues showed anomalous peaks in a difference Fourier map and side-chain distances and angles consistent with two disulfide bridges (C_β–C_β–S_γ–S_γ) (Table 3).

The structure of 9°N-7 pol reveals features common to all DNA pol structures as well as those that may be unique to archaeal pols. The overall shape of the enzyme can be described as a claw with a central hole that is lobed into NH$_2$-terminal, 3′-5′ exonuclease, and polymerase domains (Figure 1(a) and (b)). Like all other pols of known structure, the pol domain resembles a right hand and may be further divided into palm, fingers, and thumb subdomains, as was originally described for the large fragment of *E. coli* pol I (Klenow fragment) (Ollis et al., 1985). 9°N-7 pol is similar in structure to the pol α family polymerase from the mesophile bacteriophage RB69 (RB69 pol) (Wang et al., 1997), although a number of those subdomains are shorter than in RB69 pol (Figure 1(c)). Nearly all those sequence length differences are attributable to loop segments that are fewer and shorter in the hyperthermostable 9°N-7. As was first observed in the RB69 pol structure (Wang et al., 1997), the 3′-5′ exonuclease domain lies on the opposite side of the palm in comparison to pol I family polymerases. This domain arrangement is also seen in 9°N-7 pol and in Tgo pol (Hopfner et al., 1999), indicating that this motif is likely to be general for the pol α family. The structural similarity between 9°N-7 and RB69 pols is significant given the low sequence identity (<20%) in all but the active-site (palm) region, where sequence identity is 42% (Figure 2). Similar results hold for sequence alignments between 9°N-7 and human pol δ.

NH$_2$-terminal domain

Many of the members of the pol α polymerase family, including archaeal pols, thermophage T4 and RB69 DNA pols, have an NH$_2$-terminal domain that is not observed in the pol I family. T4 pol is known to control its synthesis in vivo by a mechanism of autogenous regulation (Tuerk *et al.*, 1990). The mRNA-binding activity has been located to within the first 105 residues of the pol (Wang *et al.*, 1996), but the structure of a fragment comprising residues 1–288 of T4 pol failed to suggest a structural basis for RNA binding (Wang *et al.*, 1996). Here, we note that certain structural similarities between the homologous regions in the 9°N-7 pol and the DNA RNA-binding protein may provide a reference for RNA binding by T4 pol.

The NH$_2$-terminal domain of 9°N-7 pol can be considered in three modules based on comparisons of folding (Figure 3(a)). The first module comprises residues 1–51, a three-stranded β-sheet that interacts extensively with the 3′-5′ exonuclease domain via predominantly electrostatic interactions. Residues 52–56 act as a flexible linker connecting the first module to the second (residues 57–128). The third module comprises residues 336–376.

The second module is folded into a β-β-β-β motif, with two short β-strands, 5 and 6, inserted between the second and third elements. This motif occurs in a variety of proteins, and forms the basis for the most prevalent RNA binding motif, the RNA recognition motif (RRM). The RRM is present in the RNA-binding domains of hnRNP-A1, spliceosome protein U1A and U2B″, and the sex lethal protein (Burd & Dreyfuss, 1994). Although an alignment of the NH$_2$-terminal domains of archaeal pols (Figure 3(b)), together with T4 and RB69 pols, shows that they lack the RNP1 and RNP2 sequence motifs that characterize the RRM (Burd & Dreyfuss, 1994), a number of highly conserved and consistent residue interchanges emerges. Most of these residues fall in a cluster on the surface of the NH$_2$-terminal domains of 9°N-7 and RB69 pols which therefore could mark the location of an RNA binding site atop the β-sheet platform on the face away from helix A (Figure 3(c)).

Both a sequence alignment (Figure 3(b)) and a structural comparison (Figure 3(c)) reveal that T4 and RB69 pols lack helix A and strand 2 of the β-β-β-β motif, perhaps explaining why no suggestive structural homologues to RNA-binding folds could be identified (Wang *et al.*, 1996, 1997).

Experiments are needed to determine whether the NH$_2$-terminal domain of 9°N-7 pol binds RNA. Although the β-β-β-β motif occurs in proteins that are not thought to interact with RNA (Burd & Dreyfuss, 1994), we find its presence in the NH$_2$-terminal domain of 9°N-7 pol, in a region known to bind RNA in T4 pol (Wang *et al.*, 1996), to be highly suggestive of this. RNA binding capability could hold for other archaeal pols as well, since sequence alignment of NH$_2$-terminal domain (Figure 3(b)) suggests that they share the β-β-β-β motif.

We further speculate that pols as T4 pol binds its mRNA to down-regulate its own synthesis, such autogenous regulation of pol expression might occur in archaea. Autogenous gene regulation is well documented in bacteria, and has at least one precedent in archaea. It has been identified in the synthesis of the RNA1 ribosomal protein in *Methanococcus vannielii* (Hasenbos *et al.*, 1990), and postulated for a ribosomal gene cluster from the halophile *Natronobacterium salinarium* (Shimmin & Dennis, 1990). It is interesting that there is no structural evidence that such regulation extends to eukaryotes, as human pol α shows no significant sequence homology to the NH$_2$-terminal sequences aligned in Figure 3(b).

3′-5′ Exonuclease domain

This domain is responsible for binding single-stranded DNA and excising mismatched bases in the elongated primer strand. The structure

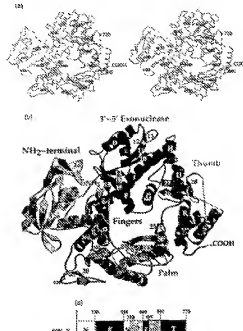


Figure 1. Structure of the Thermococcus sp. 9°N-7 DNA polymerase. The NH2-terminal and 3′-5′ exonuclease domains are colored yellow and green, respectively. The polymerase domain is divided into palm (brown), thumb (red), and fingers (blue) subdomains. Three highly conserved carboxylate groups (Dx88, Ex089, Dx5) mark the polymerase active site. (a) Stereoview of the Cα-trace. Every 50th Cα is numbered. Broken lines indicate disordered regions of the protein. (b) Ribbon diagram with secondary structure elements defined according to RB69 (Kohlstaedt & Steitz, 1992). Fifth exonuclease domain: 1, 1–30; 2, 10–22; 3, 33–35; 4, 37–49; A, 50–51; 5, 53–58; 6, 61–64; 7, 67–76; 8, 78–85; B, 92–121; 9, 109–113; C, 116–125; 1, 142–144; E, 204–206; 2–7 exonuclease-like domain; 10, 135–146; 11, 159–162; 12, 168–172; 13, 180–185; D, 187–201; 14, 205–208; F, 218–225; 15, 240–244; 16, 247–251; 17, 256–264; 9, 266–266; G, 273–285; 18, 293–305; I, 308–327. Polymerase domain: L, 329–376; 19, 397–404; M, 405–415; 20, 435–453; 10, 460–462; N, 468–469; O, 473–496; P, 507–522; 21, 535–536; 22, 563–567; Q, 585–587; 23, 595–600; 24, 610–619; 25, 625–626; R, 627–635; S, 636–651; 11, 654–665 (superimposed); T, 685–690 (superimposed); 26, 692–695; U, 697–705; 27, 709–713; 28, 714–716; V, 725–734; W, 742–746. (c) Schematic comparing the fivefold domains of Thermococcus sp. 9°N-7 and bacteriophage RB69 DNA polymerases. The domain boundaries for 9°N-7 pol were determined based upon a structure-based sequence alignment with RB69 pol (Figure 7) as defined for the RB69 pol (Wang et al., 1997).

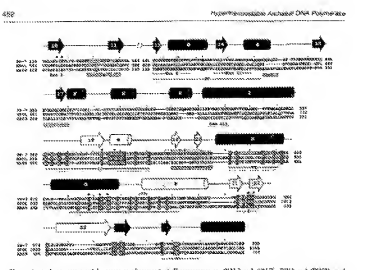


Figure 2. A thirteen-way partial sequence alignment of *Thermococcus* sp. 9°N-7 pol (9N7), RB69 pol (RB69), and human pol α (HUM). Dashes indicate gaps in the alignment, and segments not aligned are represented as twelve amino acid residue spans within brackets. Ticks mark every 50 species. The 9°N-7 and RB69 pol alignment is based upon the crystal structures. The 4RM4, and 3D6E sequences is from Wang *et al.* (1997), except for a few short segments realigned based upon the three sequences shown here. Indicated below the sequences, and boxed in yellow are conserved residues in the exonuclease (Bernad *et al.*, 1989) and polymerase (Wang *et al.*, 1997) domains. The secondary structure elements in 9°N-7 pol, as defined by DSSP, are given above the sequences. The structural elements are labeled according to the scheme described in regard to Figure 1. Shown in purple is the 9°N-7 pol sequence are the archaeal polymerase motifs described by Edgell *et al.* (1997). Residues within the polymerase domain that are conserved in the three sequences blue boxes; residues discovered in the section on dNTP binding, blue asterisks. The two disulfide bridges in the palm (CA38-C41, C74-C477) are shown (thermostable).

reported here is that of a mutant of 9°N-7 pol lacking 3'-exonuclease activity which was engineered to prevent degradation of DNA substrates during subsequent co-crystallization experiments. This 9°N-7exo⁻ pol was obtained by making two point mutations, E143A, E143A) in the first 3'-5' exonuclease domain of many DNA pols (Derbyshire *et al.*, 1986; Bernad *et al.*, 1989). In the Klenow fragment (KF) of *E. coli* DNA pol I, these residues (D368, E357) are responsible for binding the catalytic metals and for hydrogen bonding with the 3'-OH of the terminal nucleotide residue of the substrate DNA (Beese & Steitz, 1995).

Aside from loop segments that are shorter than those observed in RB69 pol (see below), the topology of the exonuclease domain in 9°N-7 pol is very similar to that of RB69 pol. The domains superimpose in the central β-sheet, containing the active site, with a root mean square deviation (rmsd) of 0.98 Å (85 C$^\alpha$ atoms). The metal binding residues not mutated in 9°N-7exo⁻ pol, D215 and D218, superimpose almost exactly on the corresponding RB69 pol residues (D222, D207).

It is now possible to assign a structural context to the four archaeal sequence motifs identified by Edgell *et al.* (1997). Three of the regions (A-C) lie within the exonuclease domain (Figure 3). Motif A forms part of the central β-sheet containing the active site; B, part of a solvent-exposed loop; and C, part of a five-stranded β-sheet nearly perpendicular to the central β-sheet. The fourth motif resides in the palm (see below).

Pol domain

This domain is responsible for the template-directed polymerization of dNTPs onto the growing primer strand of duplex DNA. Like other polymerases of known structure, the pol domain can be further divided into palm, fingers, and thumb subdomains. While the structure of the fourth of 9°N-7 and RB69 pols are highly similar, differences arise in the palm and fingers. Some of these differ-

USSN 10/089,027
Atty. Docket No. 2007651-0001

cases correspond to features that appear unique to archaeal pols, while others support a hypothesis that a conformational change occurs in the fingers as part of the catalytic cycle.

Palm subdomain

The palm, which contains the active site for polymerization, shows a high degree of structural similarity to the palm subdomain of other DNA polymerases. It is as structurally similar to pol I family polymerases as to those of the pol α family in one dimension from RB69 pol around the active site (blue region in Figure 4(b)) is 0.86 Å (95 C^{α} atoms). Together with the Tgo pol structure (Hopfner et al., 1999), this structure confirms for archaea the conservation of a common catalytic core. A significant difference between the palm subdomains in 9°N-7 and RB69 pols are the two disulfide bridges present in 9°N-7 pol, one joining Cys426 and 442 and another joining Cys504 and 509 (Figure 4(b)). Both the shortened loops and at least one disulfide bridge appear common to archaeal pols (see above). Indeed, the region containing one of the Cys residues in a disulfide bridge (C442) corresponds to the highly conserved archaeal motif G (Edgell et al., 1997; Figure 2). The Tgo pol structure shows the corresponding Cys residues to be "poised" for disulfide formation, but still in reduced form.

Until recently it was believed that all pols share a catalytic "triad" of carboxylate residues in the active site in the palm (Delarue et al., 1990). Wang et al. (1997) also recognized that only two of the carboxylate residues are invariant. The invariant carboxylates in 9°N-7 pol are D405 and D542. The third member of the triad, present as D580 in 9°N-7 pol, is not essential: mutations of the corresponding residue (D1082N) in human pol α retain catalytic function (Copeland et al., 1992). D580 in 9°N-7 pol may nevertheless be involved in binding the divalent metals required for catalysis. Mg^{2+} is normally the optimal metal for human pol α activity. The pol α D1082N mutant shows greater catalytic efficiency and fidelity with Mn^{2+} rather than Mg^{2+} (Copeland & Wang, 1993).

D580 in 9°N-7 pol interacts with the hydroxyl group of Y538 that is within hydrogen-bonding distance to D580. Substitution of this residue in Pfu or human pol α (Y526G) causes only minor effects on catalysis but alters the pol α metal affinity akin to the pol α D1082 mutation (Copeland & Wang, 1993). It seems likely that the hydroxyl moiety of Y538 in 9°N-7 pol helps to fix D580 in position for Mg^{2+}-specified binding. Consistent with this function is the strict conservation of Y538 among pol α family members (Braithwaite & Ito, 1993).

Fingers subdomain

The fingers subdomain of 9°N-7 differs in topology and relative conformation from RB69. The fingers of 9°N-7 pol use a simple hairpin-coil-helix, as in Tgo pol (Hopfner et al., 1999), whereas in the fingers of RB69 pol, the coil region is expanded with more secondary structure elements (Figures 2 and 3). The shorter fingers of 9°N-7 pol are conserved among the archaeal pols aligned in Edgell et al. (1997). It is possible that the fingers of archaeal pols define a minimal functional unit.

Different positions of the fingers subdomain relative to the palm are observed in the 9°N-7 and RB69 pol structures (Figure 5(a)). The fingers of Tgo pol (Hopfner et al., 1999) show a position intermediate between that in 9°N-7 and RB69 pols, when the palm subdomains of all three enzymes are aligned. It is interesting to note that the fingers subdomain of polymerases in the pol I family adopt different positions during the catalytic cycle (reviewed by Doublié et al., 1999). An open position corresponds to that seen in the apoenzyme form (Ollis et al., 1988; Kim et al., 1995; Kiefer et al., 1998; Kiefer et al., 1997) and the form bound to duplex DNA (Kim et al., 1995; Kiefer et al., 1998). A closed conformation has been observed in the ternary replication complexes of bacteriophage T7 pol (Doublié et al., 1998) and Klentaq (Li et al., 1998) with bound DNA and dNTP. An analogous conformational change has been observed in inverse transcriptase (Huang et al., 1998) and pol β (Pelletier et al., 1994). In the closed conformation the fingers rotate towards the palm to form a binding pocket for dNTPs.

The differences in position of the fingers subdomain in the three pol α family crystal structures suggest that the fingers of pol α family pols move during catalysis, analogous to that observed for the other polymerase families. It is interesting to note that if this is the case, there must be a corresponding enrichment in the positions of the 9°N machinery that are required in the other polymerase families as will be discussed below in the position of the fingers. In 9°N-7 pol some closely approximates a closed conformation. It is not clear why they would adopt a position previously observed only in ternary complexes with bound dNTP and DNA. The fingers of 9°N-7 pol may be stabilized in this conformation because of a salt-bridge between R309 in the palm and R480 on helix O of the fingers. These residues are highly conserved among archaeal pols (Edgell et al., 1997) and both pol I and pols α families (Braithwaite & Ito, 1993). The corresponding salt-bridge does not form in polymerases of the pol I family because the fingers helix O lies too far from the palm. The fingers of Tgo pol, as fact, are rotated slightly away from the active site, relative to 9°N-7 pol, such that the E509-R487 salt bridge cannot form. Another possible explanation for the difference in finger positions are the disulfide bridges present in 9°N-7 pol but absent in the Tgo pol structure and in pol I family structures. At least one of the disulfides (Cys426-442) in 9°N-7 pol could be directly involved in orienting the fingers relative to the palm (Hopfner et al., 1999).

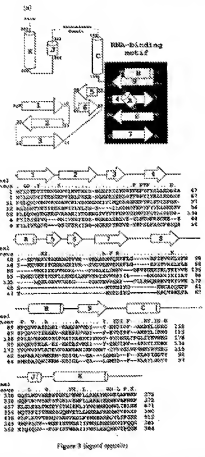


Figure 3 (legend opposite)

Model for DNA and dNTP binding

Based on the high degree of structural homology of the palm subdomain between 9°N-7 and pol I, locally pole, DNA and dNTP substrates from the bacteriophage T7 pol ternary complex (Doublie et al., 1998) were modeled into the 9°N-7 pol active site. The model shown in Figure 4 provides further

Figure 2. The RNA-binding motif in the NH$_2$-terminal domain of 9°N-7 pol. (a) Topology diagram of the complete NH$_2$-terminal domain (residues 1–428, 9°N-7). The RNA-binding motif (ribbon) located at the DNA-recognition motif (Bohr & Dvorkin, 1994), as formed, the sequence alignment of the NH$_2$-terminal domains of 9°N-7 pol, RB69 pol, T4 pol, and archaeal pols. Alignments of 9°N-7 pol and TA/RB69 is based upon the crystal structures, and that of 9°N-7 pol and the other archaeal polymerases is based upon sequence alignment of 15 sequences among twice presented by Edgell et al. (1997) (data not shown). The archaeal polymerase alignment was performed with the PILEUP algorithm at the GCG package (University of Wisconsin Genetic Computer Group). Secondary structure elements corresponding to 9°N-7 pol are given above the sequence. A consensus sequence was derived for the archaeal polymerases at those positions where at least 75% of the 15 sequences showed the same residue. Boxed in yellow are those residues conserved between the archaeal consensus and both bacteriophage (T4, RB69) sequences. Position 367 on 9°N-7 pol is starred (see the text for discussion). Abbreviations are as follows: PFU, *Pyrococcus furiosus*; SSC20, *Sulfolobus solfataricus*; VDAV, *Methanococcus jannaschii*; PDCC, *Pyrodictium occultum*; B1–B3, Ribbon1–3 representation of the NH$_2$-terminal domain of 9°N-7 (left) and RB69 (right) pols. Least-squares C$^\alpha$ superposition was performed over the region of 9°N-7 pol including strand 4, part of strand 5, and helices B and C, and the domains were separated for subtlety-side comparison. Shown in green is the putative RNA-binding motif. Charged and hydrophilic exposed connection residues are shown with green side-chains, and yellow side-chains correspond to the residues boxed in yellow in (a). The loop between β strands 3 and 4 in 9°N-7 pol corresponds to the conformationally variable loop 3 in the canonical RNP motif (Bohr et al., 1997).

evidence that the position of the fingers in 9°N-7 pol exists closely approximates a closed conformation and their position in RB69 pol approximates an open conformation. This model of a loosely complete for a pol is readily performance places the dNTP within hydrogen-bonding distance of residues on the fingers O helix that are highly conserved and known by mutagenesis to be functionally important. The corresponding residues on fingers helix P of the RB69 pol are further away and cannot directly interact with dNTP.

The model places residues Y438 and Y384 near the deoxyribose moiety of the incoming dNTP. These residues appear to be functionally analogous to F580 and Y526 of T7 pol, which are responsible for discriminating between deoxy- and ribonucleotides (rNTPs). Y438 is invariant among the pol α family. In the alignment by Braithwaite & Ito (1993) and nearly invariant (one exception) among archaeal pols aligned by Edgell et al. (1997). Mutation of the corresponding residue (Y412) to V41 to an exonuclease-deficient *Thermococcus litoralis* (Vent) pol causes a 320-fold loss of discrimination against rNTPs. The observation appears to be the functionally important moiety, as mutating Y412 in the consensus wild-type discrimination levels (Gardner & Jack, 1999).

Y526 in T7 pol (F762 in Klenow fragment) has been dubbed the "ribose selectivity site" (Tabor & Richardson, 1998). A Tyr residue at this position confers selectivity against incorporation of dideoxynucleotides (ddNTPs), whereas a Trp residue in this position allows efficient incorporation of both nucleotide species. The presence of Tyr (Y439) in this position in 9°N-7 pol suggests the ability to incorporate dideoxynucleotides, as do Vent (Gardner & Jack, 1999) and human pol α (Copeland et al., 1992). In fact, Tyr is invariant at this position among the archaeal pols aligned by Edgell et al. (1997), and highly conserved in the pol α family aligned by Braithwaite & Ito (1993).

The model of a ternary complex with dNTP and DNA places residues N480 and K481 in hydrogen-bonding distance from the triphosphate moiety of

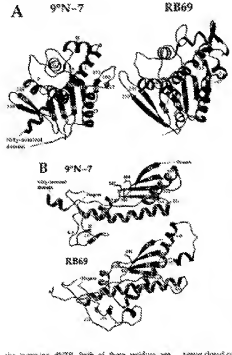


Figure 4. Comparison of 9°N-7 and RB69 pols in different functional states to indicate large segments that are shorter in 9°N-7 pol. Lettering reports C superposition was performed over the region in blue, and the domains were separated for ease-by-side comparison. Loop regions are shown in magenta and their relative endpoints are labeled. (a) Comparison of the exonuclease domains. Indicated with purple asterisks are the active site carboxylates mutated to Ala in the case of the 9°N-7exo- pol used in the study. (b) Comparison of the palm domains. The three active-site carboxylate groups are depicted with side-chains.

the incoming dNTP. Both of these residues are conserved in the pol α family (Braithwaite & Ito, 1993), and nearly identical from exemplars among archaeal pols (Edgell et al., 1997). Mutation of the corresponding residues (V486, K488 in Vent (exo-) pol severely decreases enzyme activity (Gardner & Jack, 1999).

Concerted domain movement

The shiftsense in position of the fingers subdomains in 9°N-7 and RB69 pols is part of a larger conformational change involving the 3′-5′ exonuclease and NH_2-terminal domains. Comparing these two pol structures shows that in one of the pols, an essentially rigid-body rotation has occurred involving three of the five polymerases. This concerted movement affects both the position of the fingers relative to the pol active site (open versus closed conformations) as well as the position of the exonuclease active site relative to the pol active site. The 9°N-7 and RB69 pol structures map approximate different states along the reaction pathway corresponding to DNA synthesis and 3′-5′ exonucleolytic proofreading activities.

When these two polymerases are aligned in the palm (the blue region in Figure 4b), the exonuclease and fingers are displaced between the proteins (Figure 5a). If the enzymes are aligned in the exonuclease domain (see Figure 4a), the fingers superimpose almost exactly (Figure 5b). Aligning these 3 pairs to an exonuclease-based alignment also brings the first module (residues 1–81) of the NH_2-terminal domains into identical positions (not shown). The joint evolution of the first NH_2-terminal module and the exonuclease may reflect the need to maintain ionic networks at the interface. There are two five-membered ionic net-

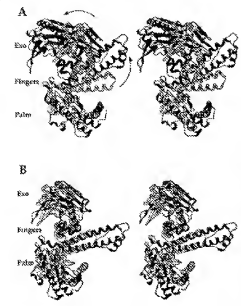


Figure 5. Least-squares C^{α} superpositions of 9°N-7 and RB69 pols in the (A) palm subdomain or the exonuclease domain. The 9°N-7 pol backbone is shown in yellow, and its active-site carboxylate groups in gold. The RB69 pol backbone is shown in green, and its active-site residues in magenta. The catalytic Mg^{2+} of the exonuclease domain is light blue (9°N-7 pol) or dark blue (RB69 pol) to allow tracking of the domain motion. The proofer regions used in the palm and exonuclease superpositions are shown in Figure 6. The NH_2-terminal domain has been omitted for clarity. Arrows in (a) indicate the direction of fingers and exonuclease movement when moving from (a) to (b).

written formed between the first module and exonuclease (Figure 7). In addition, a three-membered network is formed between the third NH_2-module (R586) and the exonuclease (Figure 7). This network is conserved among nearly all archaeal pols (Edgell *et al.*, 1997), but more so present in RB69 pol.

Comparisons of the Tgo pol structure (Hopfner *et al.*, 1999) with that of 9°N-7 and RB69 pols using palm and exonuclease-bound superpositions gives results similar to those in Figure 5, providing further support for the notion of a conserved domain movement.

A model was constructed for the RB69 pol (Wang *et al.*, 1997) showing how substrate DNA could shuttle between the pol and exonuclease active sites. When 9°N-7 and RB69 pols are aligned in the palm, the exonuclease active site in the former is filled out and away from the pol active site, making it compatible for the DNA to shuttle. The exonuclease position in RB69, but not that in 9°N-7 pol, is therefore consistent with an editing conformation. It is interesting that this creative motion also ensures that the fingers are not in position to bind dNTP (see above). Taken together, these considerations suggest that during the replication cycle of family B pols, there is concerted movement of the exonuclease, NH_2-terminal domains, and fingers relative to the catalytic region of the pols.

This concerted movement may be the structural basis for the functional coupling of polymerase and exonuclease domains, which is unique to the pol α family. In this family it is possible to generate site-directed mutations in one domain that exert an indirect, negative effect on the other (Bebo-Kinney & Nelson, 1998; Abdus-Sattar *et al.*, 1994). This contrasts with pol I pols like KF, where these activities are completely confined to their respective domains (Ollis *et al.*, 1985).

Molecular basis of thermostability

Thermococcus sp. 9°N-7 grows at temperatures of 65–90°C, and its pol has a temperature optimum of 70–80°C (Perler *et al.*, 1996). It has a half-life of 6-7

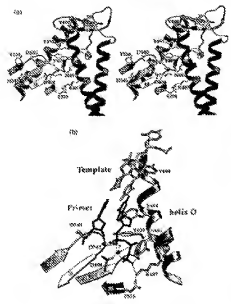


Figure 6. The active site of 9°N-7 pol and a modelled ternary complex for incorporation of the active site. Structures until indicated individual are discussed in the text. Hydrogen bonds are as broken lines, and the two disulfide bridges are shown as broken lines, and several ions are modelled as green spheres. The 9°N-7 pol and T7 pol ternary complex (Doublié et al., 1998) were superimposed on the palm (0.35 Å rmsd for 13 Cα atoms). The catalytic conformation was adopted for D542 and D654 in 9°N-7 pol, and the β turn including D542 was tilted downward, as in motion complexes in that observed between the open/apo and binary complex structures of Bacillus stearothermophilus pol (Kiefer et al., 1998).

houses at 95°C (R.B. Roberts, unpublished results), whereas Thermus aquaticus (Taq) DNA pol has a half-life of 1.6 hours at 95°C (Perler et al., 1996). The structure of 9°N-7 pol indicates a few key strategies for this hyperthermostability, some of which appear general to archaeal DNA pols.

A surprising feature of the 9°N-7 pol is that it contains two disulfide bridges (Figures 3(a) and 4(b)). The potential for the same bridges to form was also observed in Tgo pol (Hopfner et al., 1999). Although not normally the case in Bacteria or Eucarya, an increasing number of cysteinic proteins with disulfide bridges are being discovered in the Archaea (DeDecker et al., 1996; Stephens et al., 1998). The stabilizing role of disulfide bridges has been well documented (Kelehan et al., 1994; Cowan et al., 1992). Introduction of disulfide bridges therefore appears to be a common strategy for archaeal protein stability.

Alignment of a large number of archaeal pols (Edgell et al., 1997) suggests that having at least one of these disulfides at temperature for their thermostability. In fact, the two-stranded β-sheet containing C482 corresponds to sequence motif D in archaeal pols (Edgell et al., 1997). Based on structure Cys is present in the corresponding positions, all the pols observed by Edgell et al. (1997) are predicted to have at least one of the two disulfide bridges seen in 9°N-7 pol, with the exception of A1 pollur and B. stearothermophilus pol. The unavailability of A1 pollur pol may be partly caused by a lack of disulfide bridges. The B. stearothermophilus pol, like the 9 archaeal pol, is highly divergent in sequence from other archaeal pols, and it is unclear whether either of these functions as pols (Edgell et al., 1997).

An increased number of salt-bridges relative to mesostable homologs is often cited as a determinant of protein thermostability (DeDecker et al., 1996; Karshikoff et al., 1995; Chan et al., 1998; Hennig et al., 1995). The 9°N-7 pol shows a substantial increase in the fraction of charged residues participating in salt-bridges (47 %) compared with B1969 pol (30 %). These results are similar to a previous stability study of Pyrococcus furiosus glutamate dehydrogenase (Yip et al., 1995). The authors of

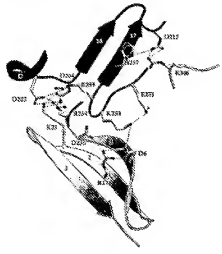


Figure 7. The extensive salt networks at the interface of the NH_2-terminal and 3'-5' exonuclease domains.

that study found a marked preference for Arg residues in the ionic interactions of the thermostable enzyme, but no such preference is evident here. The same fraction (48 %) of Arg residues is used in ionic interactions in both 9°N-7 and B86 pols, whereas a much higher proportion of Glu residues participate in salt bridges in the 9°N-7 pol (83 %) compared with B86 pol (33 %).

The number and distribution of salt bridges within thermozyme does not substantially differ between 9°N-7 and B86 pols. At the monomer–monomer (subdomains, however, the differences in ionic networks are striking. The proportion of ionic interactions at interfaces in the 9°N-7 pol (21 %) is over twice that in B86 pol (9 %). The differences lie at the interface of the exonuclease domain with the NH_2-terminal domain (Figure 7), and at the interface of the exonuclease with the thumb, where a two and a three-membered ionic network occur in 9°N-7 pol compared with none in B86 pol (not shown).

Burial of the charged groups of proteins has been cited as another factor that can confer thermostability (Hwang et al., 1995). The NH_2-terminal methione (M1) of 9°N-7 pol is stabilized by a hydrophobic cluster formed by L135, F327, I356, V358, and L141 while the corresponding residue of B86 pol is completely exposed to solvent. The β-factor for the C^{α} of M1 in 9°N-7 pol is 26 Å^2, whereas for M1 in B86 pol, it is 85 Å^2. While burial of the N terminus may be important for the thermostability of the 9°N-7 pol, the same does not hold for the C terminus. The last 25 residues are not visible in the electron density, similar to the case of B86 pol. The solvent accessibility of the C terminus of these pols may reflect the need for this region to interact with a processivity accessory protein, which is known to be the case in the T4 replication complex (Berdis et al., 1998).

Another common strategy for protein thermostability is to lower the solvent accessible surface area of the protein and to increase the proportion of buried structure (Kumchelwa et al., 1995, Chan et al., 1995). This translates into a more compact structural design. There are at least 11 examples of loop sequences in B86 pol that are much shorter or absent in 9°N-7 pol. Some of the more striking examples are shown in Figure 8. Alignment of 14 archaeal pols (Delagoi et al., 1997) indicates that they share practically all of these sequence "deletions". The Tgo pol structure also revealed shortened loop segments relative to B86 pol (Hopfner et al., 1999). Nevertheless, the overall ratio of solvent-accessible surface area to volume for both 9°N-7 and B86 pols is the same (0.33). Thus, while lowering the surface area to volume ratio is a common strategy for thermostability, it is not the primary basis for the stability of 9°N-7 pol.

Materials and Methods

Purification, crystallization, and data collection

Thermococcus sp. 9°N-7 polymerase (wild-type and the D141A,D143A exonuclease-deficient mutant) was overexpressed and purified as described (Southworth *et al.*, 1999). Crystallization, cryoprotection, data collection, and reduction of native crystals are described (Zhao *et al.*, 1998). Derivatives were prepared by soaking native crystals in stabilization solutions (Zhao *et al.*, 1998) supplemented with 22.7 mM sodium ethylenediaminetetraacetate (Sigma-Aldrich) for 11 days (thymine-O2), 10 mM K_2PtCl_4 for one hour (PtCl-1), 1.5 mM dicyclohexylammonium-(ethylenediamine)-oxoplatinate (PtP) for one day (PtP-1), or 1.0 mM Baker's isomorphic for 30 hours (BaCl$_2$). These crystals were mapped through multiwavelength anomalous scattering (MAS) data collection. 10% (five minutes), and 20% acetone (one to five hours). Additional derivatives were collected with the improved cryoprotection procedure reported (Zhao *et al.*, 1998) by soaking native crystals in 2.0 mM thymidine for 6.5 days (thymine-O2). 3.0 mM K_2PtCl_4 for seven days ($PtCl_4$), and 1.5 mM PtP for 30 hours (PtP-2).

Structure determination

The structure of the D141A,D143A mutant of 9°N-7 polymerase was determined by the method of multiple isomorphous replacement (MIR). A number of native and derivative crystals were used to solve the structure because of problems with anisotropy/twinning (Table 1). Three native datasets were collected from single crystals: NAT-1 was mounted in the frozen nitrogen stream directly from cryoprotectant, whereas NAT-2 and -3 were flash-frozen in liquid nitrogen prior to mounting. The crystals belong to space group $P6_422$, with unit cell dimensions of approximately $a = 96.1$ Å, $b = 96.1$ Å, $c = 113.3$ Å. The NAT-3 data set is the one selected to possess no asymmetric unit, giving a solvent content of approximately 46%.

A difference Patterson map of thymine-O3 was calculated using the program PP in the CCP4 suite (CCP4, 1994). One heavy-atom site for this derivative was identified with the program RSPS (Knight, 1989). This site was used to calculate initial phases for NAT-1 at 5 Å resolution using the program MLPHARE (Otwinowski, 1991). Difference Fourier synthesis with the initial phases revealed three sites for the PtCl-1 derivative. Two new sites for this derivative were determined with the phases thereof from both thymine-O 1 and PtCl-1. The correct handedness of the phasing information from these derivatives was determined using MLPHARE, and anomalous scattering data from the derivatives were included in the phase calculation. These sites for the BaCl$_2$ derivative and two other for PtP-1 were obtained from difference Fourier maps calculated to 5 Å resolution. All of these heavy-atom sites were included in subsequent phase calculations with MLPHARE. The high-resolution phasing limit was extended to 2.5 Å. Because of the high solvent content in the crystals, use of the solvent-flattening program DM (Cowtan, 1994), in combination with histogram matching, improved the phases substantially. A polyalanine model was built into the improved electron density map of NAT-1 with the program O (Jones & Kjeldgaard, 1997) and refined in the program XPLOR (Brünger, 1992). Phase combination using the program SIGMAA (Read, 1986) further improved the map during building and refinement.

Identification of side-chain densities was possible only after collecting a higher-resolution native dataset (NAT-2), along with difference data for three more derivatives obtained under improved cryoprotection conditions (thymine-O2, new data; PtCl-2, four sites; PtP-2, five sites). These derivatives were used to calculate MIR phases on NAT-2 to 3.0 Å resolution. Partial model phases of NAT-2 were calculated using the refined polyalanine model derived from NAT-1. Because of significant differences in unit cell dimensions between NAT-1 and 2, it was not necessary to subject NAT-2 to rigid-body refinement against NAT-1 on X-PLOR. Combination of the polyalanine model phases and MIR phases with SIGMAA improved the electron density map. Model building, refinement, and phase combination were extended until a complete polyalanine model could be built for the final stage of refinement. NAT-3 was used to extend the resolution limit to 2.1 Å, and water molecules were added.

Coordinate files and Illustrations

The *Thermococcus* sp. 9°N-7 polymerase atomic coordinates and structure factors have been deposited in the PDB Protein Data Bank under the accession code 3LDP. The RMS coordinates used for comparisons in this manuscript are those of the orthorhombic crystal form (coordinates under 1TAQ). Figures were prepared within the RIBBONS program (Nielsen Graphics, Inc.) entirely (1B), 7, 8(a) and 9(b) or with images imported from MOLSCRIPT (1B,c) (Kraulis, 1991) or SETOR (Chi, 4-7) (Evans, 1993).

Acknowledgments

We thank New England Biolabs, Inc. for their collaboration on this project especially S. Ray Williams and Rebecca Kucera for protein purification and Francine Perler and William Jack for helpful discussions. We thank Abbie McCoy, Dave Owen, Daniella Stack, Jean-Luc Jestin, and Hermann Heftinger for critical comments on the manuscript. We also thank Abbie McCoy, Jeff Stybar, and Sean Johnson for assistance in figure preparation. This work was supported by grants to L.S.B from the ACS (RPG-97-74, the Merck Canadian Biotechnology Career, and the Searle Scholars Program.

References

Abelson, Sutter A. K., Lin, T. C., Jones, C. & Kumagawa, M. N. (1990). Functional correspondence and mutagenic kinetic parameters of prion mutations in bacteriophage T4 DNA polymerase. *Biochemistry*, **30**, 10817–10825.

Bessa, S. M., Ollivecha, C. K., Petersen, J. D. & Pace, M. N. (1994a). Perspectives on archaeal diversity, thermophily, and monophily from environmental rRNA sequences. *Proc. Natl. Acad. Sci. USA*, **91**, 9188–9193.

Bessa, L. S. & Stetter, T. A. (1993). Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *EMBO J.* **12**, 2353.

Bessa, A. J., Soumillion, P. & Berkhout, S. J. (1990). The carboxyl terminus of the bacteriophage T4 DNA

USSN 10/089,027
Atty. Docket No. 2007651-0001

polymerase is required for heteroenzyme complex formation. *Proc Natl Acad Sci USA*, 89, 12613-12617.
Bambara, L., Berrod, A. & Sabo, M. (1992). Evidence favouring the hypothesis of a conserved 3'-5' unwinding activity site in DNA-dependent DNA polymerases. *Gene*, 123, 139-144.
Bernstein, D. S. & Ito, J. (1993). Comparative alignments and phylogenetic relationships of DNA polymerases. *Nucl Acids Res*, 21, 987-992.
Bridges, J. F. (1992). *X-Ray Structure Analysis of Some Selected Proteins and Nucleic Acids*. Yale University Press, New Haven, CT.
Burt, C. J., Wilde, G., Chase, G. J., Gross, L., Fleischmann, R. D., Sutton, G. G., Hobbs, J. A., FitzGerald, S. M., Clayton, R. A., Cocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D. et al. (1995). Complete genome sequence of the mycoplasma antelope pathogen *Mycoplasma agalactiae* sequence. *Science*, 270, 1655-1673.
Bruenger, C. A. & Steitz, T. A. (1995). Structural and functional insights provided by crystal structures of DNA polymerases and their substrate. *Curr Opin Struct Biol*, 5, 36-44.
Burd, C. C. & Dreyfuss, G. (1994). Conserved structures and diversity in functions of RNA-binding proteins. *Science*, 265, 615-621.
Chen, M. K., Mohamed, S., Kerstin, A., Adams, M. W. W. & Rees, D. C. (1995). Structure of a hyperthermophilic beta-complex enzyme, ribulose-1,5-bisphosphate carboxylase. *Science* 267, 1463-1469.
Collaborative Computational Project No. 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr*, D 50, 760-763.
Cooper, A., Eyles, S. J., Radford, S. E. & Dobson, C. M. (1992). Thermodynamic consequences of the removal of a disulphide bond: bon lysozyme. *J Mol Biol*, 225, 939-943.
Copeland, W. C., Lam, N. K. & Wang, T. S.-F. (1993). Probing residues of the human DNA polymerase α: the most conserved region among α-like DNA polymerases is not specific for nucleotide-bound pyrimidine on DNA synthesis. *J Biol Chem* 268, 17001-17009.
Copeland, W. C. & Wang, T. S.-F. (1993). Mutational analysis of the human DNA polymerase α: the most conserved region in α-like DNA polymerases is involved in metal-specific catalysis. *J Biol Chem*, 268, 11028-11040.
Crowther, R. (1984) "CCP4": an automated procedure for phase improvement by density modification. Joint CCP4 and EMBL-Hamburg Symp. Protein Crystallog. 31, 34-36.
DeDecker, B. S., Ellerson, R., Fleming, P. J., Gogol, J. H., Jackson, S. P. & Sigler, P. B. (1996). The crystal structure of a hyperthermophilic archaeal TATA-box binding protein. *J Mol Biol* 264, 1072-1084.
Delarue, M., Poch, O., Tordo, N., Moras, D. & Argos, P. (1990). An attempt to unify the structure of polymerases. *Protein Eng*, 3, 461-467.
DeLong, E. F., Wu, K. Y., Prezelin, B. B. & Jovine, R. V. M. (1994). High abundance of archaea in Antarctic marine picoplankton. *Nature*, 371, 695-697.
Derbyshire, V., Freemont, P. S. & Joyce, C. M. (1992). Structure-function analysis of 3'-5' exonuclease of DNA polymerases. *Methods Enzymol* 262, 363-385.
Doublié, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature*, 391, 251-258.
Doublié, S., Sawaya, M. R. & Ellenberger, T. (1999). An open and closed case for all polymerases. *Structure*, 7, R31-R35.
Edgell, D. R. & Doolittle, W. F. (1997). Archaea and the origin(s) of DNA replication proteins. *Cell*, 89, 995-998.
Edgell, D. R., Klaser, B.-P. & Doolittle, W. F. (1997). Gene duplications in evolution of archaeal family B DNA polymerases. *J Bacteriol* 179, 3032-3040.
Eom, S. H., Wang, J. & Steitz, T. A. (1996). Structure of Taq polymerase with DNA at the polymerase active site. *Nature*, 382, 278-281.
Evans, J. C. (1994). HERON: hardware-lighted three-dimensional solid model representations of macromolecules. *J Mol Graphics*, 12, 134-138.
Ewald, W. & Pinsent, W. (1976). Cytochromes, proteins, and measurement of subunit of 2,2'-dithiodiethanol-cytochrome in Chinese hamster ovary cells. *Cancer Res*, 36, 1369-1374.
Gardner, A. F. & Jack, W. E. (1999). Determinants of nucleotide sugar recognition in an archaeon DNA polymerase. *Nucl Acids Res*, 27, 2545-2553.
Goulian, H. S., Aggarwala, B., Francis, V. L., Sams, D. V. & Baldwin, P. (1996). Thermal stabilization of recombinant synthase by engineering two disulfide bridges across the linear interface. *J Mol Biol* 264, 69-84.
Hübner, M., Meyer, C., Rübner, G., Calvisbon, G., Cosimo, P. & Pirodi, W. (1994) Antigenous accelerated regulation of the cytosolic Wnt-1 cancer in the orthobacterium *Escherichia coli*. *J Bacteriol* 176, 609-616.
Hübner, M., Dortmund, S., Steiner, R., Kirschner, K. & Jaenicke, L. R. (1998). 2.0 Å structure of indole-3-glycerolphosphate synthase from the hyperthermophile *Sulfolobus solfataricus*: possible determinants of protein stability. *Structure*, 6, 1745-1756.
Hopfner, K.-P., Eichinger, A., Engh, R. A., Laue, F., Ankenbauer, W., Huber, R. & Angerer, B. (1999). Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*. *Proc Natl Acad Sci USA*, 96, 3600-3605.
Hübner, H., Chopra, R., Verdine, G. L. & Harrison, S. C. (1992). Structure of a transiently footprint catalytic complex of HIV-1 reverse transcriptase and implications for drug resistance. *Science*, 256, 1670-1675.
Huang, P. & Plunkett, W. (1991). Fludarabine and gemcitabine-induced apoptosis: incorporation of analogs into DNA is a critical event. *Cancer Chemother Pharmacol* 28, 181-188.
Jones, T. A. & Kjeldgaard, M. (1997). *O Version 5.9*. The Advanced Upgrade. University, Uppsala, Sweden.
Joyce, C. M. & Steitz, T. A. (1994). Function and Structure relationships in DNA-polymerases. *Annu Rev Biochem*, 63, 777-822.
Kabsch, W. & Sander, C. (1983). On the active of protein secondary structure pattern recognition of hydrogen bonded and geometrical features. *Biopolymers*, 22 2577-2637.
Kaslow, M. J., McCarthy, R. B., Bailey, G. P., Smith, T. L., Garland, R. B. & Fawcett, B. J. (1987). Impaired prospects for long-term survival in adults with acute myelogenous leukemia. *J Am Med Assoc* 258, 2681-2686.
Kiefer, J. R., Mao, C., Hansen, C. J., Basehore, S. L., Hogrefe, H. H., Braman, J. C. & Beese, L. S. (1998). Crystal structure of a thermostable Bacillus DNA polymerase I large fragment at 2.1 Å resolution. *Structure*, 6, 95-108.

Kaster, L. K., Mars, C., Brennan, J. C. & Reese, C. C. (1988) Unwinding DNA replication in a moderately extreme thermophile DNA polymerase crystal. Nature, 334, 529–307

Kim, Y., Eom, S. H., Wang, J., Lee, D. S., Suh, S. W. & Steitz, T. A. (1992) Crystal structure of Thermus aquaticus DNA polymerase. Nature 376, 612–616

Korty, Bs., Brennan, R. D. & Jack, W. E. (1993) Three-dimensional structure of a DNA polymerase from the hyperthermophilic archaeon Sulfolobus shibatae. J. Biol. Chem. 268, 19465–19473

Korolchuk, S., Hindges, R., Hübner, H., Doerschug, A. & Lanzendörfer, B. (1993) The crystal structure of bacteriophosphodiesterase I from Escherichia coli reveals the hyperthermophilic exonuclease from Thermococcus sp. 9°N-7 and mutational effects. J. Mol. Biol. 227, 857–871

Kornberg, S., Fuller, M. T., Bonner, W. M., Di Capua, E. & Thalhammer, L. (1980) Crystal structure of the large fragment of Thermus aquaticus DNA polymerase I at 2.5 Å resolution. Proc. Natl. Acad. Sci. USA, 92, 9224–9228

Li, Y., Korolev, S. & Waksman, G. (1998) Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation. EMBO J. 17, 7514–7525

Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. & Steitz, T. A. (1985) Structure of large fragment of Escherichia coli DNA polymerase I complexed with dTMP. Nature, 313, 762–766

Otwinowski, Z. (1991) In Isomorphous Replacement and Anomalous Scattering (Wolf, W., Evans, P. R. & Leslie, A. G. W., eds), pp. 80–86. Science and Engineering Research Council, Warrington, UK

Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. & Kraut, J. (1994) Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. Science, 264, 1891–1903

Perler, F. B., Kumar, S. & Kong, H. (1996) Thermostable DNA polymerases. Advan. Protein Chem. 48, 377–435

Polesky, J. F. (1992) A program to produce both detailed and schematic drawings for protein structures. J. Appl. Crystallog. 24, 548–550

Rossi, R. S. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. Acta Crystallog. sect. A, 42, 349–360

Rossi, D. C. & Adams, M. W. W. (1992) Hyperthermophiles taking the heat and loving it. Structure 3, 331–334

Rothschild, L. J. & Nancy, R. L. (1993), Genetic and biochemical studies of bacteriophage T4 DNA polymerase 3'-5' exonuclease activity. J. Biol. Chem. 36, 27010–27018

Rubenstein, L. E & Hazelbeck, W. (1992), High-dose cytosine arabinoside in chronic lymphocytic leukemia: clinical and pharmacologic synopsis. Leuk. Lymphoma, 10, 43–48

Shamoo, Y., Krueger, U., Rice, L. M., Williams, K. R. & Steitz, T. A. (1997) Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution. Nature Struct. Biol. 4, 215–222

Sharma, C. C. & Benson, P. P. (1989), Characterization of the L11, L1, L10, and L12 equivalent ribosomal protein gene cluster of the halophilic archaebacterium Halobacterium cutirubrum. EMBO J. 8, 1525–1535

Singleton, M. R., Scaglio, M. N. & Lindahl, T. A. (1998) A new structure of pyrrolidine anchored pyridine from the hyperthermophilic archaeon Thermococcus stetteri. Structure, 6, 239–244

Southworth, M. W., Kong, H., Kucera, R. B., Ware, J., Jannasch, H. W. & Perler, F. B. (1996) Cloning of thermostable DNA polymerases from hyperthermophilic marine Archaea with emphasis on Thermococcus sp. 9°N-7 and mutations affecting 3'-5' exonuclease activity. Proc. Natl. Acad. Sci. USA, 93, 5281–5285

Tahirov, T. & Richardson, C. C. (1995), A single residue in DNA polymerases of the Escherichia coli DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. Proc. Natl. Acad. Sci. USA, 92, 6209–6549

Terzi, C., Reddy, S., Pacana, D. & Cohn, L. (1994), Autogenetic translational repressor recognized by bacteriophage T4 DNA polymerase. J. Mol. Biol. 243, 789–800

Wang, J., Yu, P., Lin, T. C., Konigsberg, W. H. & Steitz, T. A. (1996), Crystal structures of an NH2-terminal fragment of T4 DNA polymerase and its complexes with single-stranded DNA and with divalent metal ions. Biochemistry, 35, 8110–8119

Wang, J., Sattar, A. K. M. A., Wang, C. C., Karam, J. D., Konigsberg, W. H. & Steitz, T. A. (1997), Crystal structure of a pol α family replication DNA polymerase from bacteriophage RB69. Cell, 89, 1087–1099

Wheat, C. R., Sandle, O. & Wilhelm, M. L. (1996), Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl Acad. Sci. USA, 87, 4576–4579

Wang, S. W., Wood, A. F., Yang, F. M., Auld, N., Thomas, B. E., Aras, R., Scott, D., Thorbjarnardo, M. W. & Wang, T. S.-F. (1998), Human DNA polymerase α gene expression is cell proliferation dependent and its promoter structure is similar to both prokaryotic and eukaryotic replicative DNA polymerases. EMBO J. 7, 37–47

Yeo, K. S. P., Sullivan, T. J., Dorian, K. L., Arrysiska, P. J., Bailey, F. L., Sudderkamp, S. E., Bogel, P. C., Zaugra, A., Chamberlain, R., Cornwell, F., Sutcliffe, S. & Rich, D. W. (1995) The structure of Pyrococcus furiosus glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. Structure 3, 1747–1758

Zhou, M., Mou, C., Rodriguez, A. C., Kirby, J. R., Pacana, R. & Steitz, L. S. (1998) Crystallization and preliminary diffraction analysis of a hyperthermostable DNA polymerase from a Thermococcus archaeon. Acta Crystallog. sect. D, 54, 994–995.

Edited by D. Rees

(Received 23 June 1999; received in revised form 24 March 2000; accepted 24 March 2000)

THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 2000 by The American Society for Biochemistry and Molecular Biology, Inc.
Vol. 275, No. 12, Issue of March 24, pp. 8523–8531, 2000
Printed in U.S.A.

Comparative Kinetics of Nucleotide Analog Incorporation by Vent DNA Polymerase*

Received for publication, July 28, 2000, and in revised form, December 21, 2000
Published, JBC Papers in Press, December 26, 2000, DOI 10.1074/jbc.M006356200

Andrew F. Gardner‡, Catherine M. Joyce§, and William E. Jack§¶

From ‡New England Biolabs Inc., Beverly, Massachusetts 01915 and the §Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511

Comparative kinetic and structural analyses of a variety of polymerases have revealed both common and divergent elements of nucleotide discrimination. Although the parameters for dNTP incorporation by the hyperthermophilic archaeal Family B Vent DNA polymerase are similar to those previously derived for Family A and B DNA polymerases, parameters for analog incorporation reveal alternative strategies for discrimination by this enzyme. Discrimination against ribonucleotides was characterized by a decrease in the affinity of NTP binding and a lower rate of phosphoryl transfer, whereas discrimination against ddNTPs was almost exclusively due to a slower rate of phosphodiester bond formation. Unlike Family A DNA polymerases, incorporation of 9-(2-hydroxyethoxymethyl)-dGTP (ganciclovir triphosphate, GTP) is efficient, cytosine, guanine, or thymine-arabinofuranosyl-TP by Vent DNA polymerase was enhanced over ddNTPs via a 16-fold increase in phosphoryl transfer rate. Furthermore, a mutant with increased processivity for nucleotide analog incorporation (Vent$^{exo^-}$ DNA polymerase) had unaltered dNTP incorporation while displaying enhanced nucleotide analog binding affinity and rates of phosphoryl transfer. Based on kinetic data and available structural information from other DNA polymerases, we propose active site models for dNTP, ddNTP, and acyclNTP selection by hyperthermophilic archaeal DNA polymerases to rationalize structural and functional differences between polymerases.

All free living organisms encode several DNA polymerases that are jointly responsible for the replication and maintenance of their genomes, thereby ensuring accurate transmission of genetic information (1–3). The majority of identified DNA polymerases can be classified into Families A, B, C, and Y according to amino acid sequence similarities to *Escherichia coli* polymerases I, II, III, and IV/V respectively (4, 5). Additional families have been identified, including the low-related replicative DNA polymerases from hyperthermophilic Archaea (Family D) (6) and eukaryotic DNA polymerases α and terminal transferases (Family X) (4).

Structural and kinetic analyses of Family A (7–14) and Family B (15–20) DNA polymerases have increased the understanding of nucleotide selection and incorporation mechanisms. Although amino acid sequences diverge between these two families, the structures of Family A and B DNA polymerases share recognizable fingers, thumb, and palm subdomains. Also, these comparisons of structural elements important for function (7, 15). In the case of Family A DNA polymerases from bacteriophage T7, *Bacteriella coli* (Klenow fragment), large fragment of DNA polymerase I, and *Thermus aquaticus*, as well as the Family B DNA polymerases from bacteriophage RB69 interpretation of the structural information is complemented by nucleotide and pyrophosphate kinetic studies, allowing a detailed description of the polymerization pathway. Reaction parameters describing the discrimination against naturally occurring nucleotide analogs encountered in cells, such as NTPs, or antiviral nucleotide analogs, such as ddNTPs and dideoxyinosine triphosphate (ddITP) (12, 21–23), have added insights into tools for nucleotide discrimination.

Hyperthermophilic archaeal DNA polymerases have not been scrutinized in such detail, hampering a complete thermodynamic and comparison with other polymerases. Family B DNA polymerases from hyperthermophilic Archaea *Thermococcus* sp. 9°N (24), *Thermococcus gorgonarius* (18), and *Pyrococcus furiosus* (strain Vent (exo$^-$)) (25) and archaeal/bacteriophage RB69 (20) have high sequence and structural homologies and provide a framework for analysis of active site structure and function in this enzyme family (Fig. 1). Furthermore, considerable biochemical studies have identified hyperthermophilic DNA polymerases catalase activities for polymerization and nuclease activities and for nucleotide binding (18, 26, 27–35). Nucleotide analogs have also been important in identifying dNTP recognition determinants. Important in the polymerase reaction (32–36) and have proven useful in a variety of molecular biology applications, such as DNA sequencing and detection of single nucleotide polymorphisms (37–41). This group of analogs, 2′-(3′)-deoxynucleoside-5′-triphosphate (dNTP) analogues (where X is adenine, cytosine, guanine, or thymine; acyclNTPs),[1] is particularly intriguing due to the wide spectrum of incorporation efficiency noted in different DNA polymerases, even within the same family of polymerases. For example, while Family B archaeal polymerases incorporate acyclNTPs more efficiently than dNTPs, eukaryotic Family B polymerase α acts as a poor substrate for acyclNTPs (42). Such differences have been exploited in drug therapies where inhibitors against specific polymerases that more readily incorporate acyclNTP than does the host DNA polymerase (43). Hyperthermophilic archaeal DNA poly-

[1] The abbreviations used are: acyclNTP, 9-(2′-hydroxyethoxymethyl)-nucleotide, where X is adenine, cytosine, guanine, or thymine; acyclGTP, 9-(2′-hydroxyethoxymethyl)-guanine triphosphate; dCTP-S, 2′-deoxycytidine-5′-O-(1-thiotriphosphate); ddTTP-S, 2′,3′-dideoxythymidine-5′-O-(1-thiotriphosphate); ddNTP, dideoxynucleotide; TAPS, 3-{[tris(hydroxymethyl)methyl]amino}-1-propanesulfonic acid.

* This work was supported by National Institutes of Health Grant GM-29832. This paper is available on line at http://www.jbc.org

¶ To whom correspondence should be addressed: New England Biolabs Inc., 32 Tozer Rd., Beverly, MA 01915. Tel.: 978-927-5054; Fax: 978-921-1350; E-mail: jack@neb.com.

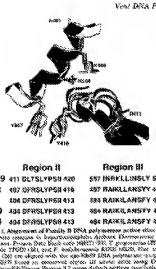


FIG. 2. Assignment of Family B DNA polymerase active sites A, ribbon trace structure of bacteriophage RB69 DNA polymerase (PDB accession code 1IG9) (Shamoo et al., 1995). The polymerase (DNA template-primer and incoming dNTP not shown) is oriented with the exonuclease domain to the left and the polymerase domain to the right. Cα atoms are depicted with the exception of Asp 368 (DNA polymerase active site) and Glu 478 (exonuclease active site) whose entire side chains are shown. Cα atoms are depicted as spheres and are labeled with the one-letter amino acid code and residue number. Figure was generated using the program Molscript (Kraulis, 1991) and rendered with Raster3D (Merritt et al., 1994). Color figure can be viewed at www.apsnet.org/publications/jbc/. B, sequence comparison of Region II and Region III motifs from family B DNA polymerases. Conserved amino acids are boxed. Numbers refer to the first amino acid in the sequence. Sequences are from RB69 DNA polymerase (accession number P16611), Vent DNA polymerase (accession number P30317), 9°N DNA polymerase (accession number U89899), KOD DNA polymerase (accession number P77933), and TGO DNA polymerase (accession number P52027). Sequences were aligned using the program Clustal W (Thompson et al., 1994). Color figure can be viewed at www.apsnet.org/publications/jbc/.

exonuclease (Vent^{R}, Deep Vent^{TM}, 9°N^{TM}, and Pfu) and exonuclease deficient (Vent^{R} (exo^{-}), Deep Vent^{TM} (exo^{-}), and Pfu) DNA polymerases with a primer/template substrate (see RESULTS) in contrast with the behavior of Taq and Klenow fragment DNA polymerases, which prefer ddNTPs (38, 44).

Recent reports of purified thermostable DNA polymerases from hyperthermophilic Archaea (Thermococcus litoralis (Vent) DNA polymerase), we identified a mutant (Vent^{R}(exo^{-})) DNA polymerase (Vent mutant) with enhanced ability to incorporate dideoxynucleotides (ddI, ddT). Subsequent crystal structures of closely related DNA polymerases strongly suggested that this residue (Ala 485 in another closely related DNA polymerase from Thermococcus 9°N-7) (Southworth et al., 1996) is located in the nascent base pair binding pocket (Fig. 1B). The nonconservative substitution of Tyr for Ala 485 in the Vent mutant DNA polymerase was later confirmed by homologous mutations in other hyperthermophilic DNA polymerases (9°N, A485Y DNA polymerase (Southworth et al., 1996); Pfu, P410Y DNA polymerase (Evans et al., 1999); and KOD, A485Y DNA polymerase (Matsumoto et al., 1998)). Further conservation of a conserved role for this residue.

Although biochemical, these steady-state observations failed to address the underlying kinetic mechanism responsible for nucleotide and nucleotide analog incorporation in hyperthermophilic DNA polymerases. Therefore, we initiated presteady-state kinetic studies to compare the modes of nucleotide discrimination in Vent and other DNA polymerases.

EXPERIMENTAL PROCEDURES

Materials. Nucleotide Analogs, DNA Substrates, and Enzymes—All DNA polymerases used in this study are 3′ → 5′ exonuclease deficient as a result of mutation of conserved aspartic and glutamic acids to alanine at the exonuclease active site (30, 31, 45). These mutations prevent nonspecific excision of newly incorporated nucleotides or analogs. Vent and Vent^{R} (exo^{-}) DNA polymerases were purified as described previously (31), and the concentration of active enzyme was determined using a pre-steady-state active site titration assay (see below). Deep Vent^{TM} (exo^{-}) DNA polymerase (Thermococcus species GB-D) and Pfu DNA polymerase (Pyrococcus furiosus) were purchased from New England Biolabs Inc. Klenow fragment (exo^{-}) DNA polymerase (Escherichia coli) was from New England Biolabs Inc. 3′-Deoxynucleotide 5′-O-(1-thiotriphosphate)(dNTPαS) and ddNTPαS were from Amersham Biosciences. 2′,3′-Dideoxy-5′-O-(1-thiotriphosphate)-adenosine (ddATPαS) was from Trilink Biotechnologies (San Diego, CA). 5′-Carboxy-fluorescein (FAM)-labeled nucleotide analogs FAM-ddCTP and FAM-acyclicTTP were provided by Phil Buzby (PerkinElmer Life Sciences) (Fig. 2). Oligonucleotides used in reactions 2′-deoxyguanosine-5′-triphosphate-dGTP) and quenching incorporation were synthesized and purified by the Oligonucleotide Synthesis Facility at New England Biolabs with a 6-carboxyfluorescein (FAM) label on the guanine nearest its 3′-terminus: 5′-FAM-GCGTTGGCAGGTCAGCAAGTGCA-3′ (20-mer; see 5′-DCMAGGCTGCGCAGGTCGACGGATCCCCGGGTACCGAGCTCGAATTCGCCCTTGCTCACCATGT-3′ (36-mer)

Pre-steady-state kinetics. DNA was formed by mixing equimolar amounts of the above-listed 20-mer primer with the 36-mer template in annealing buffer (5 mM Tris-HCl (pH 8.2 at 25 °C), 5 mM KCl, and 0.1 mM EDTA) and heating the substrate to 95 °C for 5 min, followed by annealing for 45 min at 60 °C and then cooling for 15 min at room temperature.

Burst kinetics of full-length DNA synthesis and determination of active site concentration. Most DNA polymerases used for determining the kinetics of the nucleotide step under the appropriate reaction conditions, we empirically established the reaction following burst kinetics. Burst experiments were carried out as described below with 50 nM FAM-labeled DNA, 30 or 50 nM Vent or Vent^{R}(exo^{-}) DNA polymerase, and 0.2 mM ddGTP, ddATPαS, dTTP, or acyclicTTP (final concentrations after mixing) in 1× ThermoPol buffer (20 mM Tris-SO$_4$ (pH 8.8 at 25 °C), 10 mM (NH$_4$)$_2$SO$_4$, 2 mM MgSO$_4$, and 0.1% Triton X-100). The ionic strength of the buffer was adjusted with KCl, which is equal to the active site concentration, and the initial rate of product formation (*v*, in nM product formed per second) was calculated from the burst equation: [product] = *v*(1 − exp$^{(-k_{obs}t)}$)/k_{obs} + *vt* (46). The steady-state turnover number (k_{ss}) was determined by dividing k_{obs} by the active site concentration.

Measurement of DNA Dissociation Rates Using Exonuclease Resistance. Single turnover exonuclease resistance reactions were initiated by mixing Vent or Vent^{R}(exo^{-}) DNA polymerase (0.06 μM) and FAM-labeled DNA (0.024 μM) in 1× ThermoPol buffer together with an equal volume of nuclease or exonuclease analogs in 1× ThermoPol buffer. The reactions were allowed to proceed for various time intervals and then quenched by addition of EDTA to a final concentration of 0.25–0.50 M. Reactions in the range of 0 ms to 60 s were carried out by using a rapid quench-flow instrument (KinTek Corp., Austin, TX). Reactions between 60 s and 1 h were carried out by hand mixing. Processivity was verified using enzyme/template concentrations of 0 °C. Although the temperature is lower than the optimal reaction temperature of 72 °C (72) it is the highest temperature at which the rapid quench-flow instrument can be operated reliably. Single turnover ssDNA exonuclease by Klenow fragment DNA polymerase was carried out using 0.12 μM Klenow fragment DNA polymerase (exo^{-}) and 0.06 μM FAM-labeled DNA substrate in 1× Klenow buffer (50 mM Tris-HCl (pH 7.5) and 5 mM MgCl$_2$) with an equal volume of Vent^{R}(exo^{-}) in 1× Klenow buffer. Reactions were then incubated at 55 °C for various times and quenched essentially with EDTA at a final concentration.

Conversion of the fluorescently labeled DNA polymerase substrate to product was monitored by denaturing PAGE and fluorescence detection detection methods. Product DNA was denatured by mixing a 7.5 μl aliquot of quenched sample with 40 μl of formamide and 5.5 mM EDTA and heating at 90 °C for 2 min. Fluorescent 5′-FAM-labeled 20-mer oligonucleotide substrate and 5′-FAM-labeled 36-mer oligonucleotide product bands were then resolved by electrophoresis on a 9.5 M urea, 440 mM polyacrylamide denaturing gel running on ABI377 nucleotide sequencer (Applied Biosystems, Foster City, CA) and analyzed using GeneScan Version 2.1 software (Applied Biosystems). The fraction active was assumed for subsequent steps. To verify single nucleotide incorporation was monitored over a pair of unlabeled normal bases. Rate constants (k_{obs}) were subsequently plotted as a function of

USSN 10/089,027
Atty. Docket No. 2007651-0001

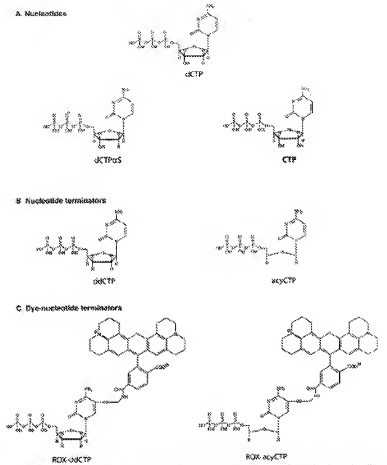


FIG. 2. Nucleotide and nucleotide analogs used for study of Vent DNA polymerase pre-steady-state kinetic mechanism. The structures of nucleotide substrate and the nucleotide analogs for binding were determined with the following nucleotides and nucleotide analogs: dCTP, dCTPαS, ddCTP, acyCTP. (A) nucleotide terminators ddCTP and acyCTP (B) acyclic 3'-hydroxyl-methyl nucleotide terminators ROX-ddCTP and ROX-acyCTP (C).

constants or analog constants above, and fitted to the hyperbolic equation $k_{obs} = k_{pol}[nucleotide]/(K_d + [nucleotide])$, yielding k_{pol}, the maximum rate of nucleotide addition, and K_d, the dissociation constant for nucleotide binding (Eq. 1). The minimum energy difference between dCTP and nucleotide analog incorporation was calculated by Equation 1 (43)

$$0.571 \times \cdots = 22 \ 100 \ k_{pol}[E_{pol}(n)]_{initial} \quad \text{(Eq. 3)}$$

Single turnover kinetics under conditions of saturating enzyme concentration. We established that in 50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 5 mM DTT, 100 μg/ml BSA, 5% glycerol, 50 mM NaCl at 72 °C, the rate of DNA strand-displacement synthesis by exonuclease-deficient Vent(exo$^-$) DNA polymerase (0.16 nM) was unaffected by the presence of 0.1% Triton X-100 (data not shown). Therefore, all subsequent experiments were performed in the presence of 0.1% Triton X-100 to prevent the adsorption of enzyme to the walls of the reaction vessel.

Measurement of Pyrophosphorolysis Catalyzed by Vent DNA Polymerase—To measure the rate of DNA strand-displacement pyrophosphorolysis catalyzed by Vent(exo$^-$) DNA polymerase, 0.16 nM enzyme was equilibrated with the DNA substrate (0.005 μM) in 1× Thermo-Pol buffer and then mixed with PP_i in 1× Thermo-Pol buffer at 60 °C using rapid quench techniques as described above. The extent of primer-strand degradation in each time point was calculated by subtracting the quantity of each DNA species in the absence of phosphodiester bonds hydrolyzed to generate that species ($f_{initial}$) and f_{max} were derived using fitting procedures described above for single-nucleotide addition.

RESULTS

Analysis of dNTP Incorporation by Vent DNA Polymerase—Previous studies with Family A DNA polymerases have shown that the steady-state pre-chemistry step for addition of a single correctly paired dNTP across phosphothioate bond linkages (Fig. 1B, 4A, 4B). Consequently, the first round of polymerization occurs more rapidly than subsequent rounds, resulting in a rapid "kick" burst of product. Incorporation of dCTP by Vent(exo$^-$) DNA polymerase displayed a burst pattern similar to those seen with RB69 exo$^-$ Asp^{222}Leu-gp43 DNA polymerase, with a rapid burst ($k_{quick} = 62\ s^{-1}$) followed by slow steady-state turnover ($k_{ss} = 0.05\ s^{-1}$) (Fig. 2A and Table I). As indicated above, the burst is diagnostic for a rate-limiting step following bond formation; moreover, the amplitude is equal to the concentration of active enzyme, indicating that ~90% of the Vent DNA polymerase preparation was active. Under similar conditions, Vent DNA polymerase failed to show a significant burst with ddCTP (Fig. 2B) or CTP (data not shown) incorporation. These data suggest that the rate-limiting step during nucleotide analog incorporation has changed compared with dNTP. Upon substitution of ddCTP with dddCTP-S, both Vent and Vent^{exo-} DNA polymerase showed a 10- and 6-fold (this alternative effect $(k_{pol(exo+)}/k_{pol(exo-)})$, respectively (Table I), increased ability to support rate-limiting step.

Determinations of K_d and k_{pol} for dCTP Addition by Vent DNA polymerase gave kinetic information similar to those determined for other DNA polymerases (Figs. 4A and Tables II and III). The relatively high K_d for nucleotide ($K_d = 70\ \mu M$ is positive on the K_d for nucleotides determined in multiple turnover steady-state measurements ($K_m = 43\ \mu M$) (31). Kinetic constants alone limit discrimination to nucleotide identity, as neither Vent DNA polymerase binding ($K_d = 26\ \mu M$) and rate ($k_{pol} = 6\ s^{-1}$) constants were observed for dATP incorporation. Substitution of dCTP with dCTP-αS had little effect on binding (K_d) or phosphodiester bond formation (k_{pol}); thus, the polymerase displays a maximum 1000-fold (k_{pol}/K_d) specificity ($k_{pol(exo+)}/k_{pol(exo-)}$) for dCTP-αS (Table II).

Analysis of Vent DNA Polymerase-catalyzed Pyrophosphorolysis—To examine Vent DNA polymerase pyrophosphorolysis activity, we employed deoxyadenosine 5′-thiotriphosphate oligomer chemistry coupled to the separation of incorporating concentrations of TP. The dependence of the rate of Vent DNA polymerase pyrophosphorolysis on PP_i concentration yielded an equilibrium dissociation constant for PP_i binding of $K_d = 340\ \mu M$ and a maximum velocity of $k_{max} = 2.1\ s^{-1}$ (Table IV).

Analysis of Ribonucleotide and Nucleotide Analog Incorporation by Vent DNA Polymerase—Kinetic parameters of ribonucleotide incorporation were determined to analyze the effect of this precursor in 2′-OH ribonucleotide on polymerization. Vent DNA polymerase discriminated strongly against CTP incorporation with a 16-fold reduced binding affinity ($K_d = 1100\ \mu M$ and 3-fold slower rate of nucleotide addition ($k_{pol} = 0.162\ s^{-1}$) (Table II). Comparisons of CTP and dCTP parameters interpreted as the ratio of relative efficiencies ($k_{pol}/K_d|_{dCTP}/k_{pol}/K_d|_{CTP}$) revealed that Vent DNA polymerase preferred dCTP over CTP by 5600-fold.

In contrast to CTP, discrimination by Vent DNA polymerase against ddCTP and acyCTP was almost exclusively due to a slower rate of nucleotide addition, with K_d values for dCTP, ddCTP, and acyCTP being roughly equal (Fig. 4B and Table II). Indeed, the approximate 20-fold preference for acyCTP over ddCTP incorporation was almost entirely be attributed to steps associated by k_{pol}.

Similar experiments with Klenow fragment DNA polymerase showed a 50,000-fold higher discrimination against acyCTP, affecting steps measured by both K_d and k_{pol}. The Klenow fragment DNA polymerase equilibrium binding constant for acyCTP was increased by 20-fold compared with dCTP and ddCTP, whereas k_{pol} for acyCTP incorporation was no-

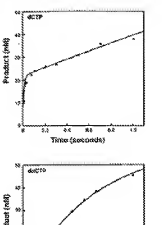


FIG. 3. Pre-steady-state burst kinetics of dCTP and ddCTP incorporation by Vent DNA polymerase. Concentrations of fluorescently labeled substrate DH (0.005 μM) were rapidly mixed with Vent DNA polymerase (0.16 nM) at 60 °C in 1× Thermo-Pol buffer and quenched with 0.3 M EDTA at the indicated times. Products were resolved from substrate by denaturing gel electrophoresis and the bands visualized as described in the legend to Fig. 2. The data were fitted to the equation $y = A(1 - \exp^{-kt}) + k_{ss}t$ by a fit. The Vent DNA polymerase dCTP burst amplitude (A) was 51 nM, the burst rate (k_{obs}) was 62 s^{-1}, and the steady-state rate (k_{ss}) was equal to 54 s^{-1}. In B, the first-order rate constant for ddCTP incorporation was 0.6 s^{-1}.

Table 1
Pre-steady-state burst kinetics

The kinetic parameters for Vent and Vent^{exo-} DNA polymerases are those at least two independent determinations and are reported as the average ± S.E. SD, not determined.

Enzyme	k_{max} (s^{-1})	$K_{max/2}$ (μM)	k_{exo} (s^{-1})	k_{exo}/k_{max}
Vent	60 ± 49	0.20 ± 0.09	0.47 ± 0.08	0.008 ± 0.002
Vent^{exo-}	80 ± 5	0.19 ± 0.01	0.38 ± 0.02	0.005 ± 0.0002
9°N^{exo-}	200 ± 99	2.2 ± 0.5	ND	ND
9°N^{exo-}/A485L	30 ± 7	2.0 ± 0.2	ND	ND

[a] Ref. 31.
[b] Ref. 19.

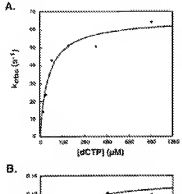


FIG. 4. Vent DNA polymerase pre-steady-state kinetics of nucleotide and nucleotide analog incorporation. The dependence of the observed rate (k_{obs}) on nucleotide or nucleotide analog concentration was fit to a hyperbola according to the Michaelis–Menten equation: $k_{obs} = (k_{max} \cdot [nucleotide])/(K_{1/2} + [nucleotide])$, where k_{max} is the observed rate of phosphodiester bond formation, and $K_{1/2}$ is the nucleotide dissociation constant. (*A*) Single-turnover kinetics of dCTP incorporation. The 3′-OH group of the primer strand was unmodified and $k_{max} = 60 \text{ s}^{-1}$. (*B*) Single-turnover kinetics of ddCTP incorporation. The 3′-OH group of the primer strand was unmodified and $k_{max} = 0.47 \text{ s}^{-1}$.

(except for ~2-fold) compared with dCTP (Table III).

ROX-ddCTP and ROX-acyclicTP Incorporation—Previous studies found ROX derivatized ddCTP and acyclicTP to be more efficient terminators than their unmodified forms when using Vent DNA polymerase (20). Pre-steady-state kinetics revealed higher binding affinities, but slower incorporation kinetics for the ROX derivatives (Table II), resulting in only marginal alterations in incorporation selectivity.

Analysis of Mechanistic Nucleotide Analog Incorporation by $9°N^{exo-}$ DNA Polymerase—We previously reported enhanced incorporation of nucleotide analogs by Vent^{exo-} DNA polymerase (18, 31). In a burst kinetics experiment, the A485L mutant enzyme gave an initial burst of dCTP incorporation at a rate similar to that seen with the wild-type enzyme ($k_{max} \approx 45 \text{ s}^{-1}$, >75% active) (Table I). Moreover, the single-turnover kinetic parameters for dCTP addition ($K_d \approx 77 \mu M$ and $k_{pol} \approx 50 \text{ s}^{-1}$) were similar to values derived for the wild-type enzyme (Table II). However, following the initial turnover, the steady-state rate of the A485L mutant polymerase was 5-fold slower than that of the wild-type enzyme ($k_{ss} = 8.39 \text{ s}^{-1}$) (Table I), accounting for the lower specific activity of Vent^{exo-} DNA polymerase (32). As with wild-type Vent DNA polymerase, replacement of dCTP with dCTPαS had little effect on K_d or k_{pol} (Table II). Vent^{exo-} DNA polymerase was less active in phosphorothioate incorporation with the wild-type enzyme, the result of a 3-fold reduction in V_f, binding affinity and a 2-fold decrease in k_{pol} (Table IV). Incorporation of ITP, dGTP, aspGTP, and ROX-ddGTP by Vent^{exo-} DNA polymerase was more efficient, by 5–10-fold compared with incorporation by the wild-type enzyme, in each case, this is attributable to both increased binding affinity (lower K_d) and faster reaction rates (k_{pol}) (Table II). Incorporation of ROX-acyclicTP was largely unaffected by the A485L mutation (Table II).

DISCUSSION

The fundamental information derived for Vent DNA polymerase incorporation of dCTP confirms and expands earlier steady-state data (31) and places Vent DNA polymerase in full consort of other Family A and B DNA polymerases. As with these other polymerases, the steady-state rate for single nucleotide addition is limited by a slow step after phosphodiester bond formation. Previous steady-state measurements using an assay in which all four dCTPs were present gave a k_{ss} value of 12 s^{-1} (31). This value is higher than the steady-state rate derived here in burst experiments (8.9 s^{-1}), most likely reflecting the higher concentrations (72 °C) and, more importantly, the processive requirements allowed by the earlier studies. In contrast, the experimental design reported here forces the DNA polymerases to act as a distributive enzyme, or, dissociating from the DNA before binding another primer-template and incorporating an other nucleotide.

The single turnover parameters for Vent DNA polymerase with the natural dCTP substrate are similar to those of other Family A and B polymerases: both nucleotide and thermodynamic. As shown in Table III, K_d and k_{pol} values define by

[31] Kong, H., M. C. B. Kucera, and W. E. Jahn, unpublished data.

TABLE II

Pre-steady-state burst constants for nucleotide and nucleoside analog incorporation by Vent and Vent$_{(exo^-)}$ DNA polymerases

In all cases, the burst phase data for Vent and Vent$_{(exo^-)}$ DNA polymerases can be fit to a two-step reversible mechanism (except where indicated). The constants are also reported in the format: $x \pm y$.

Nucleotide	Vent DNA polymerase				Vent$_{(exo^-)}$ DNA polymerase			
	K_d	k_{pol}	k_{pol}/K_d	Incorporation[a]	K_d	k_{pol}	k_{pol}/K_d	Efficiency[b]
	μM	s^{-1}	$M^{-1} s^{-1}$		μM	s^{-1}	$M^{-1} s^{-1}$	
dCTP	79 ± 7	68 ± 1	8.6×10^5		72 ± 9	57 ± 3	7.9×10^5	
ddCTP/ddA	1.0 ± 0.05	0.0033 ± 0.0001	3.3×10^3	1 A	0.8 ± 0.1	0.0026 ± 0.0	3.3×10^3	1.0
dTTP	1700 ± 200	0.0033 ± 0.0001	1.9×10^3	600	200 ± 30	0.0033 ± 0.0	1.9×10^4	6.0
ddaTTP	80 ± 5	0.18 ± 0.01	2.3×10^3	270	15 ± 0	0.05 ± 0.01	1.9×10^3	42
ddaCTP	50 ± 20	0.4 ± 1.1	2.7×10^3	32	20.5 ± 0.9	2 ± 0	5×10^4	6.4
3OH-ddaCTP	35 ± 1	0.0039 ± 0.0001	1.7×10^3	923	0.6^c	0.02^c	3.6×10^4	70
2OH-ddaCTP	0.6 ± 0.2	0.002 ± 0.001	2.5×10^3	4	0^c	0^c	1.8×10^4	4.4

[a] Refers to the factor by which incorporation of ddCTP and other NTPs is the ratio of the efficiency of dCTP incorporation (k_{pol}/K_d) to the efficiency of the indicated NTP.
[b] Refers to the factor by which incorporation of ddCTP and other NTPs is the ratio of the efficiency of dCTP incorporation (k_{pol}/K_d) to the efficiency of the indicated NTP.
[c] Estimated.

TABLE III

Pre-steady-state burst constants for nucleotide analog incorporation by DNA polymerases

The kinetic parameters for Vent and Vent$_{(exo^-)}$ DNA polymerases are from at least two independent determinations and are reported as the means ± S.D. All rate measurements were performed at 37 °C.

Enzyme	dCTP				ddCTP				3OH-ddaCTP	
	K_d	k_{pol}	k_{pol}/K_d	Incorporation[a]	K_d	k_{pol}	Efficiency	K_d	k_{pol}	Selectivity[b]
	μM	s^{-1}	$M^{-1} s^{-1}$		μM	s^{-1}		μM	s^{-1}	
Vent	79 ± 7	68 ± 1	8.6×10^5	9000	45 ± 7	0.18 ± 0.01	930	35 ± 1	0.0039 ± 0.0001	30
Vent$_{(exo^-)}$	72 ± 9	57 ± 3	7.9×10^5	450	15 ± 0	0.05 ± 0.01	60	20.5 ± 0.9	2 ± 0	3.4
D580N	80 ± 10^3	105 ± 6^3	9.76×10^5	$14,000^3$	4500 ± 900^3	0.13 ± 0.03^3	$29,000^3$	750 ± 100^3	0.2 ± 0.1^3	270^3
H589A	8.0 ± 2.0	20 ± 3^3	25 ± 7^3	$3,600^3$	650 ± 120^3	0.4 ± 0.2^3	620^3	40 ± 20	more than 6000	6000
Y409B	30 ± 3^3	7.5 ± 0.7^3	250^3	$36,000^3$	360^3	0.03 ± 0.01^3	8300^3	950^3	0.03	0.09

[a] Refers to the factor by which incorporation of ddCTP and other NTPs is the ratio of the efficiency of dCTP incorporation (k_{pol}/K_d) to the efficiency of the indicated NTP.
[b] Refers to the factor by which incorporation of ddCTP and other NTPs is the ratio of the efficiency of dCTP incorporation (k_{pol}/K_d) to the efficiency of the indicated NTP.
[c] Ref. 29.
[d] Ref. 35.
[e] Ref. 37.
[f] Ref. 19.

TABLE IV

Pyrophosphorolysis

The kinetic parameters for Vent and Vent$_{(exo^-)}$ DNA polymerases are from at least two independent determinations and are reported as the means ± S.D.

Enzyme	$K_{app(PP_i)}$	k_{pol}	$k_{pol}/K_{app(PP_i)}$	$\alpha K_{app(PP_i)}$
	μM	s^{-1}	$M^{-1} s^{-1}$	μM
Vent	160 ± 160	1.30 ± 0.07	8.2×10^3	300
Vent$_{(exo^-)}$	1000 ± 800	0.03 ± 0.01	3.7×10^3	1800
H589F	10,000	0.01	1.2×10^3	180,000
H589W	1000	0.01	2.0×10^3	6000

[a] A measurement of DNA polymerase activity with pyrophosphorolysis (pathway 3) is given by the ratio of DNA polymerase efficiency ($k_{cat}/K_{app(PP_i)}$) divided by the efficiency of pyrophosphorolysis ($k_{pol}/K_{app(PP_i)}$).
[b] Ref. 20.
[c] Ref. 9.

(H589A) for all polymerases tested, with no clear direction between Family A and B DNA polymerases. Furthermore, the Family A Klenow fragment and Family B Vent and H589 DNA polymerases carry out the reverse reaction of DNA polymerases, pyrophosphorolysis, with similar rates (k_{pol}), and Klenow fragment and Vent DNA polymerases share comparable PP$_i$ binding constants ($K_{app(PP_i)}$) (Table IV). Similarities in catalytic incorporation kinetics and active site structures underscore the evolution of DNA polymerases to efficiently carry out DNA replication and repair. Significant kinetic differences between the polymerases become apparent only when measuring nucleotide analog incorporation and aberrational effects, as detailed below.

Ribonucleotides—Despite a similar level of selectivity against NTPs, the discrimination is supplied almost exclusively by steric-gate resistance by k_{pol} for Klenow fragment DNA polymerases, whereas Vent DNA polymerase shows not only K_d effects but also a 16-fold weaker ground state binding of the nucleotide. H589 DNA polymerase also shows effects in both K_d and k_{pol}, culminating in even higher discrimination by virtue of a 300-fold weaker ground state binding. Discrimination against 3′-OH sugars, as large pools, have attributed to a steric clash between the 3′-OH and a conserved Asn clash in the polymerase active site (96, 98, 126). The kinetic parameters suggest that the steric clash is first encountered in the ground state nucleotide binding by Vent exo$^-$ H589 DNA polymerases, but does not affect Klenow fragment DNA polymerase until the transition state of the reaction. This could occur for example, if the K_d term for Klenow fragment DNA polymerase includes conformational changes prior to a rearrangement that engages the 3′-OH sensing residue(s).

Dideoxynucleotides—When incorporating ddCTP, H589 DNA polymerase discriminates at the level of both K_d and k_{pol}. Discrimination by Vent, Klenow fragment, and Kod-Taq terminated Taq DNA polymerase with a 2H nucleic acid N exocyclic deficient (2H) DNA polymerases is almost exclusively as seen

...tions increased by k_{pol} and not those increased by K_D with Yeast DNA polymerase showing less discrimination than the other two polymerases. This parallel behavior appears to reflect a similar lack of dNTP contributions to ground state substrate binding rather than a conserved set of nucleotide contacts.

On the surface, the non-linearity in k_{pol} relative to dNTP concentration by Yeast and Klenow fragment DNA polymerases may suggest similar discriminatory mechanisms. In these two enzymes, a nucleotide misaligned by the absence of an elemental subset with dNTPase acting either enzyme. The simplest interpretation is the lack of an elemental effect with ...ribonucleotide dNTPs with Klenow fragment and Yeast DNA polymerases is that a monofunctional step preceding ...phosphodiester bond formation is rate-limiting. Similarly, the lack of a significant chemistry effect for Klenow fragment DNA polymerase incorporation of dNTPs...[13] means that steps preceding phosphodiester bond formation continue to be rate-limiting for that enzyme. In contrast, the substantial effect noted for Yeast and Yeast[1995] DNA polymerase incorporation of dNTPs is an indication that the chemistry of phosphodiester bond formation significantly influences the rate-limiting step for these polymerases.

The k_{pol} ratios with both polymerases were significantly slower for rNTPs than for dNTPs. 8000- and 400-fold for Klenow fragment and Yeast DNA polymerases, respectively in the case of Yeast DNA polymerase, this must reflect at least a slowing of the chemical rate, whereas for Klenow fragment DNA polymerase, at least the rate of pre-chemical conformation is slowed. Thus, the pre-chemical rate for Yeast DNA polymerase dNTP incorporation is at least 10-fold faster than corresponding steps for Klenow fragment DNA polymerase.

Conserved amino acids positioned within active Family A or B DNA polymerases active sites probe for accurately base-paired substrates and concomitantly align phosphates into a geometry required for nucleophilic reaction. As conserved by Franklin et al. (this and Yang et al. (26) in the RB69 DNA polymerase ternary enzyme/substrate/ion complex analogy, in the Yeast DNA polymerase active sites (Fig. 1), the dNTP deoxyribose moiety maintains a favorable 3′-endo sugar conformation. This conformation is maintained by hydrogen bonds between the 3′-OH and a main chain amide (corresponding to Yeast DNA polymerase position 1128) and a non-bridging α phosphate oxygen (Fig. 5, A and B). Nucleoside α, β, and γ phosphates are further stabilized by direct or water-mediated hydrogen bonds which arrow this residues (Fig. 5, A and B). The absence of the 2′ OH on rNTPs destroys hydrogen bonding with the β phosphate and main chain amides, potentially increasing the activation energy required to orient the α phosphate for phosphoryl transfer (Fig. 5C). Indeed, the measured energetic difference between dNTP and rNTP incorporation (2.8 kcal mol^{-1}) is equivalent to this expected for the loss of at least two hydrogen bonds in the dNTP transition state (27).

Although the active site inserting network differs, in the Family A Klenow fragment DNA polymerase the dNTP 3′ OH contributes 3.7 kcal mol^{-1} to transition state stabilization, accounting for inefficient dNTP incorporation (27). This energy loss is counteracted by the closely related 3′ DNA polymerase active site by a hydroxyl group in Tyr[766] (Klenow fragment DNA polymerase has Phe in the analogous position) that interbases a hydrogen bond to stabilize the dNTP α phosphate in the transition state, re-establishing a hydrogen bonding network similar to interactions formed by dNTP 3′OH. As a result, 3′ DNA polymerase selectivity between dNTP and rdNTP is greatly reduced, as is the selectivity of the analogous Phe to Tyr mutation in both Klenow fragment and Taq DNA polymerases (28).

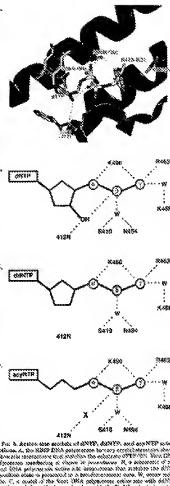


FIG. 5. Active site models of dNTP, dNTP, and acyNTP substrates. A, the RB69 DNA polymerase ternary enzyme/primer/template/dNTP ...ctions that stabilize the substrate dNTP (26). Yeast DNA polymerase numbering is shown in parentheses. B, a schematic of the Yeast DNA polymerase active site interactions that stabilize the dNTP transition state is presented in a two-dimensional view. W, water molecule. C, a model of the Yeast DNA polymerase active site with dNTP. Loss of the 2′ OH destroys hydrogen bonds with the β phosphate and dNTP 3′ OH removes hydrogen bonds with Tyr 766. Addition of the 2′ OH suggests that, on the assumption of a similar ring, a substituted 2′ OH could be a weak acceptor, but no substitute hydrogen-bonding between the Tyr[766] main chain amide and non-bridging β phosphate.

Exo(-)Vent(R)-DNA Polymerase. Divergence between polymerases is often noted upon exo(-)/3′ mutation, again suggesting divergent modulation for nucleotide recognition and incorporation between the polymerases. Similar to dNTP, exo(-)/3′ task the 3′-OH required to establish a hydrogen bonding network between the sugar, three metals of $Tp^{3/4}$, and β phosphate of the substrate (Fig. 5D). Klenow fragment-DNA polymerase displays a strong discriminatory role on both K_d and k_{pol} terms, resulting in a selectivity value of 32,000 (in this case, the efficiency of exo(-)/3′ incorporation is nearly as low as for an incorrect base pair A_{anti}:K_d = 240 and 460 s^{-1}, respectively; Table III) (16). A marginal rate against exo(-)/3′ incorporation has also been noted for Taq DNA polymerase (34).

In contrast, exo(-)/3′s are incorporated by hyperthermophilic archaeal DNA polymerases with only 10-fold lower efficiency than dNTPs by analogy with dNTP incorporation by T7 DNA polymerase. It seems reasonable that the space normally occupied by the sugar 3′- and 2′-carbons and associated substituents would be intolerable to steric confinement. Instead, the protein may have established interactions to complement for those energies lost by the missing 3′ OH group. The difference in efficiencies from energy between dNTP and exo(-)/3′ incorporation ($\Delta\Delta G^{\ddagger}$ = $\Delta G^{\ddagger}_{dNTP}$ - $\Delta G^{\ddagger}_{exo(-)/3'}$ = 4.0 kcal mol^{-1}) is equivalent to a pair of two additional hydrogen bonds, which could be provided by hydrogen bonding between the main chain 112 amide, a putative water, and an exo/TP β-phosphate nonbridging oxygen, to induce interactions that exist in the dNTP active site (Fig. 5D). At the same time, we cannot rule out stabilizing interactions arising from residues near the active site potentially excluded by the ribose 2′- and 3′-carbon that are absent in exo/TP. Clearly, three-dimensional structural analysis will be necessary for a full understanding of the interactions important for Vent DNA polymerase incorporation of exo/TPs.

Base-unstabilized Nucleotides. Base-enhanced nucleotides have been useful in a variety of analytical applications (40, 41). Not surprisingly, given the diversity of spin structures and charges, base-unstabilized nucleotides are accepted by DNA polymerases with varying efficiencies (28, 29, 35, 36). Previous studies identified nucleotide derivatives containing five-membered spin ROS as being more efficiently incorporated by Vent DNA polymerase than are pyrimidine nucleotides lacking the dye (28, 35), on the current kinetic window, the magnitude of enhanced dye-unstabilized terminator additions was much lower than previously anticipated in nonquantitative gel extension analysis, even though some base pair retention attempt was good agreement with the relative incorporation efficiency of dd/dNTP and exo/dNTP substrates (37). ROS substitutions of the nucleotide results in a 6.5–10-fold lower K_d, suggesting that contacts to or adjacent to the Vent DNA polymerase active site stabilize dye binding. However, in the same time, k_{pol} is reduced, suggesting that one consequence of the increased binding is to slow nucleotide addition. Thus, polymerase incorporation is a balance of both binding and catalysis: a substrate bound with too high affinity requires higher activation energy for efficient turnover by the polymerase.

Vent$^{exo(-)}$ DNA Polymerase Pre-steady-state Kinetics. Previous studies identified a Vent DNA polymerase mutant (A488L) with enhanced nucleotide analog incorporation properties (38, 39). Current dNTP incorporation by Vent$^{exo(-)}$ DNA polymerase is characterized by similar burst-like efficacy (E_{max}), nucleotide turnover rate (k_{pol}), and rate-limiting step consistent with Vent DNA polymerase, presumably reflecting the conservation of residues actively involved in coordinating the incoming dNTP. In contrast, each of the nucleotide analogs tested with Vent$^{exo(-)}$ DNA polymerase have higher binding affinity and faster rates of phosphoryl transfer than the unmodified polymerase. Energy differences between Vent and Vent$^{exo(-)}$ DNA polymerase incorporation of ddNTP or exo/dNTP are similar, $\Delta\Delta G^{\ddagger}_{dd/dNTP}$ = 4.8 kcal mol^{-1} and $\Delta\Delta G^{\ddagger}_{exo/dNTP}$ = 4.8 kcal mol^{-1}, suggesting that subtle hydrophobic or hydrogen bond-mediated effects could account for enhanced analog incorporation.

One hypothesis to account for these effects envisions the A488L variant no longer closes to the activated conformation, thus facilitating incorporation of analogs. The residue analog given to Ala^{488} in the RB69 DNA polymerase crystal structure points away from the active site and base at the interface between the ternary core of the polymerase and as a helix that must make a 60° rotation to form the closed complex (Fig. 1A and 1B). DNA polymerases positioning a longer leucine residue at the position normally occupied by alanine in the α-helix may shift equilibrium from the open toward the closed conformation, thus reducing the activation energy for both binding and nucleotide transfer. This comes at a price: faster exonuclease discrimination that subsequent turnover by the A488L variant is inhibited, perhaps reflecting hindrance of the translation from closest to open states required for release and/or binding of the template and dNTP. This property does not, however, easily account for the fact that pre-steady-state kinetics for the transient substrates are unaltered in this viscous environment. resolution of this discrepancy may be on the genuine ability of this variant to overcome distortions in the nucleotide-binding site, distortions that are not present when the natural nucleotide is bound.

In summary, from these comparative studies, we observed that location of dNTP incorporation pathways are conserved among Family B and B DNA polymerases despite diversity in primary amino acid sequence, macromolecular, function, and biological roles. However, differences in exo/TP and other nucleotides making catalytic efficiencies by Klenow fragment, Vent, and other DNA polymerases illustrate fundamental differences underlying the kinetic pathway for DNA polymerization. As more DNA polymerases are studied functionality is to appreciate that notable structural variations in the active site influence how nucleotides are bound and positioned for catalysis.

Acknowledgments. We are grateful to Christie Hutchinson for answering this research as part of the Master of Internal Area Program at Harvard University to A. P. G. We also thank Phil Bailey for providing the base-labeled compounds and for helpful discussions about Klenow fragment fragments, for useful discussions and technical advice, Hiroki Shibata, Satoshi Iwahashi, and Tina Brown for critical review of the manuscript, and Carla Bessett, Lucia Giacometti, and Claire Moore for providing expert technical assistance. We are indebted to Dan Tawfik for fostering a supportive research environment at New England Biolabs Inc.

REFERENCES

1. Kornberg, A. (1980) *DNA Replication*, pp. 87-97, W. H. Freeman & Co., San Francisco.
2. Joyce, C. M., and Steitz, T. A. (1994) *Annu. Rev. Biochem.* 63, 777–822
3. Doublié, S. et al. (1998) *J. Mol. Biol.* 278, 1067–1084
4. Kiefer, J. R., Mao, C., Braman, J. C., and Beese, L. S. (1998) *Nature* 391, 304–307
5. Kim, J., Prasad, C., Rao, Lee, T., and Lawrence, C. (1995) *J. Mol. Biol.* 254, 382–394
6. Kunkel, T. A., Eastman, H., Pan, H., Sabatt, S., and Loeb, L. A. (1980) *Proc. Natl. Acad. Sci. U. S. A.* 77, 7405–7409
7. Polesky, A. H., Dahlberg, M. E., Benkovic, S. J., Grindley, N. D. F., and Joyce, C. M. (1992) *J. Biol. Chem.* 267, 8417–8428
8. Steitz, T. A. (1999) *Curr. Opin. Struct. Biol.* 9, 31–38
9. Doublié, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998) *Nature* 391, 251–258
10. Doublié, S., Bandwar, R. S., and Ellenberger, T. A. (1998) *Biochemistry* 36, 6038–6050
11. Johnson, K. A., Taylor, J. S., and Brown, L. R. (1982) *Proc. Natl. Acad. Sci. U. S. A.* 79, 3656–3660
12. Mildvan, A. S., and Fersht, A. R. (1977) *Biochem. Biophys. Res. Commun.* 49,

[illegible]

USSN 10/089,027

Atty. Docket No. 2007651-0001

4567604v1

doi:10.1006/jmbi.2000.4455 available online at http://www.idealibrary.com on IDEAL J. Mol. Biol. (2001) 305, 405–417

JMB

Crystal Structure of DNA Polymerase from Hyperthermophilic Archaeon *Pyrococcus kodakaraensis* KOD1

Hiroshi Hashimoto[1], Motomu Nishioka[2], Shinsuke Fujiwara[2] Masahiro Takagi[2], Tadayuki Imanaka[3], Tsuyoshi Inoue[1] and Yasushi Kai[1]*

[1]Department of Materials Chemistry and [2]Department of Biotechnology Graduate School of Engineering, Osaka University 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

[3]Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University Yoshidahonmachi, Sakyo-ku Kyoto 606-8501, Japan

*Corresponding author

The crystal structure of family B DNA polymerase from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1 (KOD DNA polymerase) was determined. KOD DNA polymerase exhibits the highest known exonuclease rate, processivity and fidelity. We carried out the structural analysis of KOD DNA polymerase in order to clarify the mechanisms of these extraordinary features. Structural comparison of DNA polymerases from hyperthermophilic archaea highlighted the conformational differences in Thumb domains. The Thumb domain of KOD DNA polymerase shows an "opened" conformation. The fingers subdomain possessed many basic residues at the side of the polymerase active site. The residues are considered to be accessible to the incoming dNTP by electrostatic interactions. A 6-thiopurin motif (residues 242–249) extends from the Exonuclease (Exo) domain as seen in the editing complex of the RB69 DNA polymerase from bacteriophage RB69. Many aromatic residues are located at the forked-point like junction of the template-binding and editing clefts of KOD DNA polymerase, suggesting that the basic environment is suitable for partitioning of the primer and template DNA duplex and for stabilizing the partially melted DNA structure in the high-temperature environments. The stabilization of the melted DNA structure at the forked-point may be correlated with the high PCR performance of KOD DNA polymerase, which is due to low error rate, high elongation rate and processivity.

Keywords: archaea; crystal structure; family B DNA polymerase; "forked-point"; KOD DNA polymerase

Introduction

DNA polymerases are a group of enzymes that use single-stranded DNA as a template for the synthesis of the complementary DNA strand. These enzymes are multifunctional, with both synthetic (polymerase) and new or less degradative modes (3′-5′ and/or 5′-3′ exonucleases) and play an essential role in nucleic acid metabolism including the processes of DNA replication, repair and recombination. Many DNA polymerase genes have been cloned and sequenced. Amino acid sequences deduced from their nucleotide sequences can be classified into four major types. Escherichia coli DNA polymerase I (family A), *E. coli* DNA polymerase II (family B), *E. coli* DNA polymerase III (family C) and others (family X).[1] Recently, a new family of DNA polymerases has been identified, all members of this family contain five highly conserved motifs, I–V, and several of these polymerases participate in lesion bypass.[2] This family is called the Dpo4/Dbh family.[3] Family B DNA polymerases include eukaryotic DNA polymerases α, δ, and ε, which are thought to be components of the replisome and to carry out chromosomal DNA replication. Archaeal proteins involved in gene expression, such as those for DNA replication, transcription, and translation, have been found to be similar to those from eucarya. Therefore, the archaeal system of gene expressions is a simplified model of the eukaryotic system. In contrasts, the

* E-mail address of the corresponding author: kai@chem.eng.osaka-u.ac.jp

0022-2836/01/020405–13 $35.00/0

cellular appearance and organization of archaea are more like those of bacteria.

The first crystal structure of a family B DNA polymerase to be obtained was that of bacteriophage RB69 DNA polymerase (RB69 DNA polymerase).[7] The first crystal structure of archaeal DNA polymerase was DNA polymerase from *Thermatoga maritima* (*Tma* DNA polymerase).[8] The editing complex of RB69 DNA polymerase has been reported,[9] but further crystal structures of archaeal family B DNA polymerases have recently been reported: Tma DNA polymerase from *Thermococcus* sp. 9°N-7 DNA polymerase from *Thermococcus* sp. 9°N-7.[10]

The *Pyrococcus kodakaraensis* KOD1 is a hyperthermophilic archaeon, with an optimum growth temperature of 95–C.[11] Enzymes produced in KOD1 were reported to be extremely thermostable and to have eukaryotic characteristics.[12] The optimum temperature of KOD DNA polymerase is 75–C, similar to that of DNA polymerase obtained from *Pyrococcus furiosus* (Pfu DNA polymerase). KOD DNA polymerase, however, exhibits the largest extension rate (100–130 nucleotides/second) and processivity (>2 kb/cycle) five times and ten to 15 times higher than those of Pfu DNA polymerase, respectively.[13] Thermostable DNA polymerases are expected to be available enzymes for Polymerase Chain Reaction (PCR). KOD DNA polymerase is, therefore, suitable for DNA amplification by such means. Indeed, KOD DNA polymerase is widely used in rapid and accurate PCR systems (TOYOBO Ltd., Japan).

Although structures of three archaeal DNA polymerases have been determined as described above, no structural information relating to exonuclease activity, processivity or fidelity is provided. We carried out the structural analysis of KOD DNA polymerase in order to clarify the mechanism of exonucleic features of KOD DNA polymerase, which are the highest extension rate, processivity and fidelity. Here, we report the crystal structure of DNA polymerase from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. The three-dimensional structure of the KOD DNA polymerase may provide useful information to clarify the mechanisms for rapid and accurate synthesis. In addition, this information may contribute to the improvement of the PCR properties of enzymes already in use such as thermostability, error rate, elongation rate and processivity, or for designing new enzymes for PCR as well as DNA replication by family B DNA polymerases.

Results and Discussion

Overall structure

KOD DNA polymerase has a dish-like shape with dimensions of 65 Å × 65 Å × 102 Å and is made up of distinct domains and subdomains N-terminal (Ntter 1–108, 327–366, violet), Exonuclease (Exo 131–326, blue), Polymerase (Pol) domain including the Palm and Fingers subdomains (388–449, 503–607, brown, and 450–498, green, respectively) and the Thumb domain including thumb-1 and thumb-2 subdomains (368–374, red) (Figure 1(a)). The polymerase active site, containing three conserved carboxylates, Asp404, Asp606 and Asp542) is located in an anti-parallel β-sheet in the Palm subdomain. The exonuclease active site contains two conserved carboxylates (Asp141 and Glu143) and is located in an anti-parallel β-sheet in the Exo domain. The Polymerase and exonuclease active sites are the molecular surface are indicated by P and E, respectively (see Figure 4). Structural components of ordered DNA polymerases (KOD, Tgo and 9°N-7 DNA polymerase) are shown in Figure 1(b). The structural architectures of the proteins are identical, but the orientations of the domains and subdomains is different. In the case of the KOD DNA polymerase (red), the Thumb domain is shifted to make an "open" conformation and the position of the Palm domain neighboring the root of the Thumb domain is slightly shifted as a result of the large movements of the Thumb domain in comparison to other ordered DNA polymerases. Table 1 shows the averaged temperature factors of the domains and subdomains in the crystal structure of KOD DNA polymerase. The value of the Thumb domain was markedly higher than the others. The structures of many residues in the Thumb-2 subdomain are not defined, because the connection of the subdomain is highly disordered. Therefore, it is thought that the structure of KOD DNA polymerase described here provides information for the DNA-free most relaxed conformation. The structure of the editing complex of RB69 DNA polymerase revealed that newly synthesized duplex DNA is grasped by the Pol and Thumb domains. Although the orientation of the Thumb domain is presumably highly flexible, the orientation may be fixed when it binds to the primer-template duplex.

Polymerase domain

The Pol domain is made up of the Fingers and Palm subdomains and has an "L-like" shape (Figure 2(a)). The polymerization mechanism has been studied mainly on family A DNA polymerases (Pol-I). A structural basis for a motif

Table 1 Averaged temperature factors

Domain	Temperature factor (Å²)
Ntter	58.1
Exo	70.7
Pol	
Fingers	69.9
Palm	72.8
Thumb	85.7
Overall	75.9

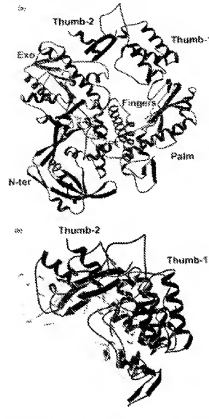


Figure 1. (a) Overall structure of KOD DNA polymerase. The structure is composed of domains and subdomains which are N-terminal, Exo, Palm, Thumb-1, Thumb-2, Fingers and C-terminal. Exonuclease (Exo) domain: Palm domain: Thumb-1 domain including the Palm (green) and Fingers (green) subdomain and the Thumb domain (red), including the Thumb-1 and Thumb-2 subdomains. Conserved carboxylate residues in Palm and exonuclease active site are shown by ball-and-stick models. (b) Conformational comparison of Thumb domains among three archaeal DNA polymerase. Red: KOD DNA polymerase; blue: Tgo DNA polymerase and green: *9°N-7* DNA polymerase. The comparison shows that the Thumb domain of KOD DNA polymerase displays the most "opened" conformation.

assisted mechanism of phosphoryl transfer was presented by the bacteriophage T7 DNA replication complex.[1] The complex structure shows that two metal ions are located in strictly conserved carboxylates (Asp475 and Asp654, which correspond to Asp501 and Asp645 in KOD DNA polymerase) extended from two anti-parallel β-sheet of the Palm domain. The phosphate group of incoming dNTP is held by the metal ions and the four basic residues extending from the Fingers subdomain (Lys486, Arg518 and Lys522). The crystal structure of two tertiary complexes of the large fragment of

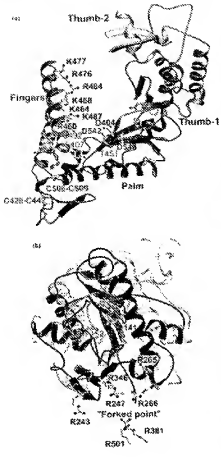


Figure 3. (a) Ribbon representation of the Pol domain. The domain is made up of Fingers and Palm subdomains. Conserved carboxylate residues (D404, D548 and E743) are concentrated by hollow-arch models. Basic residues are represented by ball-and-stick models, which stand as a lane in the Fingers subdomain, facing the polymerase active site. RNA are replaced by alkanes, because of the ambiguity of its electron density. Two disulfide bonds are displayed (C408–C442 and C505–C509). Aromatic residues adjacent to a glycine residue, represented by ball-and-stick models, are localized in the lobes of the subdomains. The Thumb domain is separated by the semitransparent screen. The C$^{\alpha}$ atoms of the methionine residues, M37 (the Pol-I), M412 (in Pol-α) and T541 (in Pol-β), are represented by violet spheres. (b) Ribbon rendered diagram of KOD DNA polymerase and other DNA polymerases. Semitransparent cylinder, Conserved carboxylate (D141 and E143) and aspartic residues (D323, D327, D405, D463, D465, D540 and D653) in the active-site of KOD DNA polymerase are represented by ball-and-stick models. Basic residues around the polymerase active site participating template-binding and substrate-binding are also represented by ball-and-stick models. The last remaining Thr587 is shown in cascade. Ile33 of RB69 DNA polymerase and P559 of KOD DNA polymerase are represented by stick-representation and opaque ball-and-stick models, respectively.

thermus aquaticus DNA polymerase I (Klentaq) with a primer-template DNA and ddCTP have been reported.[34] The ternary complexes suggest that basic residues at the fingers subdomain bind the phosphate group of the incoming dNTP and the dsDNA induces a conformational change to deliver the incoming nucleotide to the active site. In the case of family B DNA polymerases, the fingers subdomain is composed mainly of two long helices and does not have a palm that augmented in the structures of family A DNA polymerases. Therefore, it seems that in the case of archaeal DNA polymerases, the mechanism of the Fox domain to deliver dNTP to the active site differs from that of family A DNA polymerases. Kinetic studies of RB69 DNA polymerase mutants revealed that three residues (Arg482, Lys486, Lys580 and Asn584) of the Fingers subdomain affected dNTP incorporation.[35] The mutants are conserved in family B DNA polymerases, and correspond to Arg481, Lys485, Lys487 and Asn491 in 9°N DNA polymerase, respectively. Furthermore, Lys580, Arg576, Lys477 and Arg484 are located at the tip of the Fingers subdomain on the side of the polymerase active site in RB69 DNA polymerase (Figure 2(a)). It is expected that the "closure" of these residues captures the incoming dNTPs, then the dNTP is delivered toward the polymerase active-site center by accompanying the movement of the polymerase domain. Two disulfide bonds exist in the connection site between the Palm and Fingers subdomains (Figure 2(a): Cys524-Cys652 and Cys506-Cys509). The two disulfide bonds are found also in the crystal structures of Tgo Pol and 9°N-7 DNA polymerases. Sequence alignments for archaeal DNA polymerases is shown in Figure 3, suggesting the potential for the formation of disulfide bonds at the same sites. It is thought that the disulfide bonds are required to maintain the structure of the Fingers and Palm subdomains at extremely high temperatures. Sequence comparison suggests that the number of disulfide bonds are correlated with optimum growth temperatures of organisms. DNA polymerases from *Thermococcus litoralis*, *Methanococcus jannaschii* and *Archaeoglobus fulgidus*, with optimum growth temperatures at 85, 85 and 85°C, respectively, are expected to have one disulfide bond, because Cys486 is replaced by serine in *T. litoralis* and *M. jannaschii*, and Cys442 is replaced by arginine in *A. fulgidus*. DNA polymerase from *Methanobacterium thermoautotrophicum*, with an optimum growth temperature of 65°C, is expected to have no disulfide bond, because Cys438, Cys442 and Cys506 are replaced by glutamic acid, arginine and serine, respectively.

Archaeal DNA polymerases have characteristic sequences of aromatic residues adjacent to glycine residues (Figure 3). These are located at the hinges of the Palm subdomain at the connections to the Fingers and Thumb-2 subdomains (Figure 2(a)). These aromatic residues may provide a flexible structure; structure-wise, because of the adjoining glycine residues. They may contribute to the conformational changes of Fox domain in polymerization.

The 3′-5′ exonuclease domain

DNA is synthesized by competition between the rate of polymerase and exonuclease activities at the newly synthesized 3′ terminus from the primer. Misincorporation of a nucleotide destabilizes the structure of duplex DNA at the 3′ terminus of the primer. This decreases the rate of nucleophilic attack on the α-phosphate group of the incoming dNTP by the primer 3′-OH and allows excision of the incorrect nucleotide by the proofreading exonuclease. The complex requires the movement of the 3′ terminus to the exonuclease active site accompanied by rewinding of the duplex DNA, because the exonuclease active site is set apart from the polymerase active site in RB69 DNA polymerase, the exonuclease active site is set apart from the polymerase active site by approximately 40 Å. The editing complex of RB69 DNA polymerase shows structural similarity to the editing mode of family B DNA polymerase.[9] The DNA polymerase binds the mismatched primer-template DNA, which is partially denatured, the 3′ end of the primer strand is located at the exonuclease site. Residues 253–262 of RB69 DNA polymerase, that form an extended β-hairpin structure that juts directly out from the protein surface and projects into the DNA, destabilize the partially denatured or melted structure. Arg263 extending from the β-hairpin motif plays an important role. Arg263 and Phe123 appear to block the template strand by making interactions with the penultimate base at the 3′ end of the primer-template. Arg263 and Phe123 in RB69 DNA polymerase correspond to Arg287 and Phe182, in RCD DNA polymerase respectively (Figure 2(b)) shows the structural comparison of Exo domains of RCD and RB69 DNA polymerases. Molecular surface and ribbonstructure potentials are shown in Figure 4. The β-hairpin motif in RCD DNA polymerase corresponds to residues 282–289 and Arg287, extending to the forked point, which is the junction of the template-binding and editing cleft (T-cleft and E-cleft, respectively) (Figure 4c, d) means that Arg287 can maintain template strand from primer strand and stabilize the melted structure of the strands in a manner similar to that of the RB69 DNA polymerase. As Phe182 is not apart from the active site, it is apparently unable to make an aromatic interaction with the base of the primer. Based on the above idea, the movement of the loop including Phe182 (Figure 2(b)) is expected to interact with the primer bound at the E-cleft. Furthermore, Arg247 extends from the β-hairpin structure to the T-cleft. Arg247 interacts with the template strand to lie it at the T-cleft. It exhibits to Arg263 and Arg287. Five arginine residues gather at the forked-point in RCD DNA polymerase (Arg286, Arg266, Arg346, Arg383 and Arg301) and provide a basic environment (Figures 2(b) and 4). It seems that they can interact with the phosphate

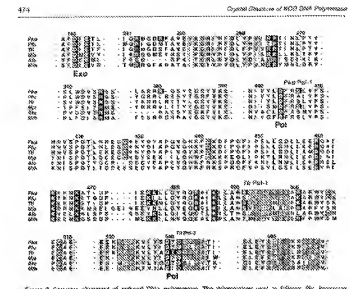


Figure 3 Sequence alignment of archaeal DNA polymerases The abbreviations used in Rfamer, Pfu, Pyrococcus furiosus; Pw, Pyrococcus horikoshii; Tli, Thermococcus litoralis; Mja, Methanococcus jannaschii; Aph, Archaeoglobus fulgidus and Mth, Methanobacterium thermoautotrophicum. Homologous residues are marked in gray. Remarkable residues are highlighted in reverse type. Conserved carboxylate residues in the Exonuclease and Polymerase active sites are shown in red. Basic residues gathering at the forked-helix and Fingers subdomains are shown in blue. R183, R207, R209, R260, R286, R491 and R561 are located at the forked-point. R460, R464, R480, R481, R483, R487, R560 and R680 are located in the Finger subdomain and face into the polymerase active site. Cysteine residues forming the possibly reversible disulfide bonds are shown in green. Hydrophobic residues in self-splicing regions are shown in violet. Introns/inteins both in the endonuclease residues. Aromatic residues adjacent to glycines are shown in orange.

groups of the DNA strand and stabilize the melted structure of DNA strands at the forked-point. Several arginine residues at the forked-point are conserved in known family B DNA polymerases from hyperthermophilic archaea.

In DNA synthesis, the structure of DNA is variable at the stage of switching between the elongation- and melting modes. Hyperthermophiles must have mechanisms to protect their genomic DNA against thermal denaturation. The genomic DNA of hyperthermophilic archaea have nucleosome-like structure brought about by interaction with histone-like proteins[14]. Nevertheless, as the replicative fork, the DNA strands are exposed. Therefore, DNA polymerases of hyperthermophilic archaea are required to stabilize the melted or melted DNA structure at the high temperature environment. The stabilization by DNA polymerase may correlate with the enzymatic characteristics of DNA polymerase such as fork-Rh period of activity, error rate, elongation rate and processivity. As discussed above, it is considered that the arginine residues around the "forked-point" have a remarkable effect on the stability of DNA structure. In the forked-point of Pfu DNA polymerase, Arg247, Arg363 and Arg501 are replaced by methionine, threonine and lysine, respectively. Therefore, the replacements may affect the difference of the enzymatic characteristics between RB69 and Pfu DNA polymerases. Additional experiments such as site-directed mutagenesis, a together with enzymatic studies of DNA polymerases are necessary to clarify the role of the residues at the forked-point.

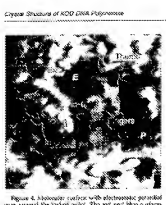


Figure 4. Molecular surface with electrostatic potential map around the bound palm. The red and blue surfaces are acidic and basic regions, respectively. Domains and subdomains are labeled with orange letters. Polymerase and thumb-base active sites are labeled with E and I, respectively. Pol B helices is labeled with β.

Exten connection site

The KOD DNA polymerase gene encodes a 3671 amino acid nucleic protease protein. The precursor protein is processed precisely into three parts by protein splicing. The self-splicing reaction yields the mature KOD polymerase (775 residues) and two intervening protein domains (internal protein, Pol-Pol-2 (360 residues) and Pol-Pol-3 (567 residues) as a result of the ligation of the external N and C-terminal domains (termed exteins).[6][7] All known precursor proteins contain conserved amino acids at self-splicing sites: serine, threonine or cysteine (hydroxyl-amino) at the intein N terminus, and His-Asn pair at the intein C terminus followed by serine, threonine, cysteine (nucleophile) at the C-extein, N-terminus.[7] The intein of the protein splicing reaction in KOD DNA polymerase are Ser407 and Ser493, which were located in at the N terminus of the C-intein. In the crystal structure of KOD DNA polymerase, the nucleophilic residues are found in the Pol-domain (Figure 6A).

Self-splicing sites in archaeal family B DNA polymerases in pyrolyti are conserved into three types: Pol-Pol-1, Pol Pol-3 and Pol Pol-2 (The Intein Database, http://www.neb.com/neb/inteins.html). The nucleophilic residues, serine or threonine, in the three sites are mapped in Figure 2(a). In the case of KOD DNA polymerase, Pol-Pol intervening in the Pol Pol-1 site and Pol-Pol3 intervening in the Pol Pol-3 site. The structure shows that they are localized around the polymerase active site in the Palm domain. Although they are exposed to solvent, they are surrounded by the fingers subdomain and the Thumb domain. The two loci(s) connect exist in the space because of steric hindrance. Therefore, it is necessary that the folding of intein and the subsequent self excision are carried out before the extein is folded.

Materials and Methods

Crystallization

KOD DNA polymerase was overexpressed in *E. coli* BL21(DE3) and purified by the previously reported method.[6] The crystals of KOD DNA polymerase were grown by the previously reported method.[17] KOD DNA polymerase was concentrated up to about an $A_{280\ nm}$ of 25. Crystals of KOD DNA polymerase suitable for diffraction experiments were obtained at 288 K with hanging drops of 2 μl of protein solution and 2 μl of reservoir solution containing 100 mM sodium citrate buffer (pH 5.6 and 25–30% (v/v) 2-methyl-2,4-pentanediol (MPD) equilibrated against the reservoir solution.

Data collection

X-ray diffraction measurements were performed at the beamline 18B of the Photon Factory at the High Energy Accelerator Research Organization, Tsukuba, Ibaraki Ken, Japan. Each crystal of KOD DNA polymerase was picked up directly with a nylon fiber loop from a drop of mother liquid, the crystal was then rapidly transferred to the N_2 gas stream. The frozen beam with wavelength of 1.00 Å was collimated to 0.2 mm at detector. Intensity data were collected on 205 mm × 420 mm imaging plates (Fuji Film Company Ltd.) using the Weissenberg camera for macromolecules with a distance of 420 mm[18] and the oscillation method with 2° oscillation per frame. The crystals diffracted to least to 2.8 Å resolution at 100 K. X-ray diffraction data were processed and scaled with programs DENZO and SCALEPACK.[19] The diffraction data were scaled with zero σ (cutoff). Unit-cell parameters were determined as $a = 131.9$ Å, $b = 127.4$ Å, and $c = 74.0$ Å, with the space group of $P2_12_12_1$. The asymmetric parameters gave Matthews' coefficient (V_m)[20] and solvent content as about 3.0 Å^3/Da and 59.8%. The final merging statistics of the data consisted of 39,587 independent of 98,288 unique observed reflections with an overall R_{sym} of 6.4% and 94.5% in the outermost resolution shell (3.96–3.40 Å). The symmetry 98.1% in theoretically measurable reflections at 3.4 Å resolution. The outermost resolution shell of data is 85.7% complete.

Structure determination

The crystal structure of KOD DNA polymerase was solved by molecular replacement with the AMoRe program.[21] The structure of the DNA polymerase (pdb code 1TGO) reduced to polyalanine was used as a search model. Data in the resolution range of 20.0–3.5 Å were used in the rotation and translation functions. Positional and temperature factor refinements of the AMoRe correlation coefficient (CC). Using a Fabiocon cut-off radius of 35 Å, a list of 25 solution families peaks was obtained, with the top peak having an AMoRe CC value of 19.8. The top solution by translation function is CC of 83.7 with an R factor of 54.1%. At this stage, the electron density of the Thumb domain is very ambiguous. Therefore, structure refinements of the initial stage was carried out with

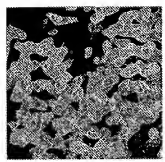


Figure 8. The final $2F_o - F_c$ map around the Phe410 and Tyr409 side-chains. The map is contoured at 1.2 σ.

a model covering the Thumb domain. The model was manually rebuilt using the program $O^{[24]}$ and subjected to further rounds of refinement using data in the resolution range 40.0–3.0 Å with the program CNS.[25] The final R-factor is 23.1% and R_{free} is 31.3% with exclusion of the bulk solvent correction. The r.m.s. deviations for bond lengths and bond angles being 0.007 Å and 1.3°, respectively. The 28 residues at the tip of one Thumb domain are not included in the final model due to poorly defined electron density. Figure 8 shows the final $2F_o - F_c$ map superimposed on the refined final coordinates of KOD DNA polymerase.

Protein Data Bank accession code

Refined coordinates and structure factors have been deposited in the RCSB Protein Data Bank under the accession code 1JCL.

Figure preparation

Figures 1 and 2 were prepared using programs MOLSCRIPT[26] and RASTER3D.[27] Figure 3 was prepared by CHORE.[28] Figure 5 was prepared using the program GRASP.[29]

Acknowledgments

We thank Professor H. Sakabe, Dr N. Watanabe, Dr M. Suzuki and Dr H. Igarashi, for support in data collection at KEK-PF, Japan. This study was supported by TARA Sakabe Project at University of Tsukuba. The author is grateful for a JSPS Fellowship for Japanese junior scientists.

References

1. Ito, J. & Braithwaite, D. K. (1991). Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucl. Acids Res.* **19**, 4045–4057.
2. Ito, J. & Braithwaite, D. K. (1990). Structure, function and evolution of DNA polymerases. *Prog. Nucl. Acid Res. Mol. Biol.* **39**, 109–143.
3. Braithwaite, D. K. & Ito, J. (1993). Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucl. Acids Res.* **21**, 787–802.
4. Wang, J., Sattar, A. K. M. A., Wang, C. C., Karam, J. D., Konigsberg, W. H. & Steitz, T. A. (1997). Crystal structure of a pol α family replication DNA polymerase from bacteriophage RB69. *Cell*, **89**, 1087–1099.
5. Hopfner, K.-P., Eichinger, A., Engh, R. A., Laue, F., Ankenbauer, W., Huber, R. & Angerer, B. (1999). Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*. *Proc. Natl Acad. Sci. USA*, **96**, 3600–3605.
6. Shamoo, Y. & Steitz, T. A. (1999). Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell*, **99**, 155–166.
7. Zhou, Y., Jeruzalmi, D., Sharma, J., Leighton, L., Lasken, R. & Kuriyan, J. (1999). Crystal structure of an archaebacterial DNA polymerase. *Structure*, **7**, 1289–1299.
8. Rodriguez, A. C., Park, H.-W., Mao, C. & Beese, L. S. (2000). Crystal structure of a pol α family DNA polymerase from the hyperthermophilic archaeon Thermococcus sp. 9°N-7. *J. Mol. Biol.* **299**, 447–462.
9. Hopfner, K. Eichinger, S. & Lipscomb, T. (1999). The mutant of archaeal aspartate semialdehyde dehydrogenase from *Thermococcus gorgonarius* at 1.75 Å. *J. Mol. Biol.* **288**, 619–632.
10. Takagi, M., Nishioka, M., Kakihara, H., Kitabayashi, M., Inoue, H., Kawakami, B., Oka, M. & Imanaka, T. (1997). Characterization of DNA polymerase from *Pyrococcus* sp. strain KOD1 and its application to PCR. *Appl. Environ. Microbiol.* **63**, 4504–4510.
11. Dauter, Z., Timmins, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature*, **394**, 181–186.
12. Li, Y., Korolev, S. & Waksman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* **17**, 7514–7525.
13. Yang, G., Franklin, M. J., Li, J., Konigsberg, W. H. (1993). Structure-function domain characterization of RB69 DNA polymerase: identification of a unique insertion. *Biochemistry*, **32**, 8295–8305.
14. Hopfner, K. P., Eichinger, A. & Rieger, J. P. (1998). DNA stability and DNA binding proteins. *Mitteil. Osterr. Chem. Ges.* **68**, 437–467.
15. Nishioka, M., Fujiwara, S., Takagi, M. & Imanaka, T. (2000). Characterization of two exonuclease-containing endonucleases associated on the KOD1 polymerase gene of *Pyrococcus kodakaraensis* strain KOD1. *Mitel. Acids Res.* **28**, 4699–4612.
16. Perler, F. B., Davis, E. J. & Ashen, E. (1992). Comparison and analysis of intein sequences. *Nucl. Acids Res.* **20**, 5087–5095.
17. Kuroshima, H., Matsumoto, T., Nishioka, M., Yasui, T., Takesada, R., Inoue, T., Fujiwara, S., Takagi, M.,

Iwamoto, T. & Kit, Y. (1998). Crystallographic studies on a family B DNA polymerase from hyperthermophilic archaeon Pyrococcus kodakaraensis KOD1. *J. Biochem.* **123**, 983–986.

18 Sakaguchi, N., Kumasaka, K., Sakaguchi, K., Higashibo, T., Nishiyama, A., Watanabe, N., Ashida, S. & Sasaki, K. (1998). Structural basis for intramolecular translocation of the chromosomal primer strand after DNA synthesis. *Mol. Cell. Biol.* **18**, 6838–6845.

19 Sakaguchi, N., Nishiyama, A., Ashida, S. & Sasaki, N. (1998). Macromolecular crystallography station BL-18B at the photon factory. *J. Synchrotron Rad.* **5**, 1024–1026.

20 Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.

21 Stuttfeld, E. & Ballmer, W. (1998). Software control of precision crystals. *J. Appl. Cryst.* **31**, 477–487.

22 Navaza, J. (1994). Amore: an automated package for molecular replacement. *Acta Crystallogr. sect. A* **50**, 157–163.

23 Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. sect. A* **47**, 110–119.

24 Bricogne, G. I., Ashton, P. D., Chow, G. M., Gilmore, W. L., Girn, P., Gomez-Susantes, R. W., Jiang, J.-S., Kantardjieff, J., Rittgers, M., Pearce, N. S., Read, E. J., Rice, L. M., Simpkins, T. & Womack, C. L. (1998). Cyclicallography & 2DMF (protein: a new software suite for macromolecular structure determination. *Acta Crystallogr. sect. D*, 54, 905–921.

25 Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950.

26 Merritt, E. A. & Murphy, M. E. (1994). Raster3D version 2.0: a program for photorealistic molecular graphics. *Acta Crystallogr. sect. D* **50**, 869–873.

27 Merritt, E. A. & Bacon, D. J. (1997). Raster3D photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524.

28 Nicholls, A., Sharp, K. A. & Honig, B. (1991). GRASP: Graphical rapid and approximate analysis of hydrophobic properties of biomolecules. *Proteins Struct. Funct. Genet.* **11**, 281–296.

Edited by R. Huber

(Received 21 August 2000; received in revised form 8 December 2000; accepted 11 December 2000)

Crystal structure of an archaebacterial DNA polymerase

Yanxiang Zhao[1†], David Jeruzalmi[1†], Ismail Moarefi[1,2], Lore Leighton[1,2], Roger Lasken[3] and John Kuriyan[1,2*]

Background: [illegible]

Results: [illegible]

Conclusions: [illegible]

Addresses: [illegible]

†The first two authors contributed equally to this work.

*Corresponding author.
E-mail: [illegible]

Key words: [illegible]

Received: 12 April 1999
[illegible]
Accepted: [illegible] June 1999

Published: 20 September 1999

Structure October 1999, 7:[illegible]

Introduction

DNA polymerases can be classified into at least three families on the basis of sequence similarities [illegible]

[illegible]

USSN 10/089,027
Atty. Docket No. 2007651-0001

4567604v1

However, the subdomain architecture of the Pol I family enzymes is the RB69 structure, even though the detailed structures of the polymerases are quite divergent [7]. The exonuclease domains of Pol I and Pol II DNA polymerases are distinct related in sequence and, not surprisingly, the structure of the exonuclease domain of RB69 resembles that of the Pol I type polymerases. Even the general architecture of the 3′–5′ exonuclease domains in RB69 was a surprise. In RB69 the 3′–5′ exonuclease domain is located above the fingers and opposite the thumb subdomain, suggesting that the shuttling of DNA between the polymerization and proofreading sites must occur by a different mechanism to Pol I DNA polymerases [7].

The mechanism of the Pol I family DNA polymerases is also undisputed (a detail [8]; Nature 28). The elongation of nucleotide addition is catalyzed by two magnesium ions that are located in the palm subdomain, at the base of a deep cleft in the polymerase domain. High-resolution crystal structures of the Pol I type DNA polymerases of *T7 bacteriophage* (T7 Pol) and *Thermus aquaticus* (Taq Pol) complexed to primer-template DNA and incoming nucleotide have been determined, revealing the mechanisms of nucleotide incorporation and selectivity to be visualized [9,11,28]. Although corresponding structural information for the Pol II family DNA polymerases is lacking, similarities in general organization of the polymerase core as well as conserved cross-motifs within crucial elements of the ternary pol/DNA substrate suggest that general features of the recognition of DNA will be similar in Pol II polymerases.

The DNA polymerase from the archaebacterium *Desulfurococcus strain* Tok (D. Tok Pol) is a member of the Pol II family and has high exonucleolytic DNA polymerase and 3′–5′ exonuclease activities [13]. D. Tok Pol contains undetectable DNA polymerase activity after incubation at 95°C for one hour [13], unpublished results. The sequence of D. Tok Pol is very closely related to 70% identity to that of other archaebacterial DNA polymerases, such as those from *Pyrococcus furiosus* [14] and *Thermococcus gorgonarius* [15]. D. Tok Pol is also related to eukaryotic DNA polymerases α, δ and ε [16], sequence identity over 30% residues of the DNA polymerase core for the human δ sequence [17]. The archaebacterial genome also contains genes coding for proteins with close homology to proliferating cell nuclear antigen (PCNA), the DNA polymerase clamp in eukaryotes, as well as subunits of the clamp-loader complex, RFC. Lastly, factor C1. It is likely that archaebacterial DNA polymerases achieve processivity by interaction of the ring-shaped PCNA sliding clamp with these enzymes but such a mechanism is lacking.

We have determined the structure of D. Tok Pol at 2.6 Å resolution. D. Tok Pol shares less than 30% sequence identity with RB69 Pol, yet the structures of the two enzymes resemble each other closely. The structure reported here was later determined in the absence of DNA. Accordingly, the close structural correspondence between the active sites of Pol I and Pol II DNA polymerases allows inferences to be made about the mode of DNA recognition by D. Tok Pol. The very N-terminal region of D. Tok Pol possesses a domain (residues 1–121) that is distantly related in structure to single-stranded DNA-binding domains (SSBDs) also known as RNA-recognition modules (RRMs) [18]. The structure of the 3′–5′ proofreading exonuclease domain of D. Tok Pol is similar to those of the Pol I type polymerases. However, as it is in relation to the palm subdomain, resembles the fingers unit in RB69 Pol rather than the Pol I type polymerases [6,10,11]. The structure of D. Tok Pol reported here provides further evidence that the probe of DNA-template recognition and the distinct editing channel established for the Pol II family by the structure of RB69 Pol is valid for the entire Pol II family.

Results and discussion

Structure determination

Crystals of D. Tok Pol have been obtained from 2-hydroxyethyldisulfide (HPD) (Sigma-D) and polyethylene glycol (PEG) 400 (Merck-H). Both crystal forms are orthorhombic $P2_12_12_1$, $a = 59.8$ Å, $b = 189.8$ Å, $c = 163.2$ Å for Native I and $a = 60.1$ Å, $b = 190.6$ Å, $c = 155.9$ Å for Native II. Experimental phases (7.6 Å) at 2.6 Å were obtained from four thimerosal heavy-atom derivatives, using Native II and the program HEAVY [19]. Phases were improved via iterative cycles of real space density modification, averaging of solvent flipping and negative density truncation, using SOLOMON [20,21]. The resulting electron-density map allowed the chain to be traced easily. Especially, with easily determination of sequence register. The model was refined to 2.6 Å against data for Native II (R value = 24.2%, R_{free} = 29.6%) and subsequently, to 2.0 Å against data for Native I (R value = 25.3%, R_{free} = 29.9%), using CNS [22]. The model for Native II is somewhat more complete than the Materials are synthetic activities and is used for most of the discussion. This model includes 790 residues from 1 to 776 in Native II. Arginine 420–426 and 635–636 are not visible in the electron-density map and are not included in the model.

General description of the structure

D. Tok Pol (Figure 1) is composed of a polymerase domain (residues 396–775) and an exonuclease domain (residues 122–385), as well as an N-terminal domain (residues 1–121) that is not found in Pol I type DNA polymerases [4]. The polymerase domain is further comprised of three smaller subdomains, termed the thumb (residues 669–776), palm (residues 760–843 and 900–959) and fingers (residues 480–659). The structure of the ΔNPD and PED480 crystal forms of D. Tok Pol are very similar

Table 1

Data collection, structure determination and refinement statistics.

	Resolution (Å)	Number of unique reflections	Completeness (%)	R_{sym} or R_{merge} (%)	R_{iso} (%)	Sites	Phasing power*	Figure of merit
Native data	–	–	–	–	–	–	–	–
Native 1	50.0–2.6	42,009	92.8(53.8)	5.2(35.6)	–	–	–	–
Native 2	50.0–2.7	40,349	99.2(90.8)	4.0(31.8)†	–	–	–	–
SeMet-analog	–	–	–	–	–	4	–	0.267
1λ	50.0–3.0	32,284	91.5(93.3)	8.4(39.9)	10.1	6	1.34/0.992	0.219
2λ	20.0–3.0	34,963	94.3(79.4)	9.5(29.6)	13.5	1	1.18/0.988	0.195
3λ	50.0–3.0	35,161	95.0(87.4)	10.2(31.5)	10.1	6	1.99/0.989	0.227

Refinement	Number of Reflections (F) > 2σ	R_{cryst}/R_{free} (%)	Final number of atoms	Protein (a) helices (Å)	Ramachan region (%)	Mean (B-value (Å²)
Native1	50.0–2.6	21.6/31	24,329	2.13/1	90.2/9.7/3	1.8/91
Native2	50.0–2.6	21.2/28	25,829	2.1/0	0.0/0.2/0.3	1.3/47.9

$R_{sym}(I) = 100 \times \Sigma(|I - \langle I \rangle|)/\Sigma I$, where I is the observed intensity and $\langle I \rangle$ is the mean intensity of reflection. $R_{iso}/R_{cryst} = 100 \times \Sigma||F_{PH}| - |F_P||/\Sigma|F_P|$, where F_{PH} and F_P are the structure-factor amplitudes for derivative and native crystals, respectively. *Phasing power = $\Sigma|F_H|/\Sigma||F_{obs}| - |F_{calc}||$, where $|F_H|$ is the calculated heavy-atom structure factor amplitude. †Figure of merit = $\langle|\Sigma P_i(\alpha)e^{i\alpha}/\Sigma P_i(\alpha)|\rangle$, where α is the phase and $P(\alpha)$ is the phase probability distribution. ‡Ramachandran plot statistics: favored (%)/allowed (%)/outliers (%). §$R_{free} = 100 \times \Sigma|F_{obs} - F_{calc}|/\Sigma|F_{obs}|$, calculated using 10% of the data. Numbers in parentheses apply to the highest resolution shell.

in terms of the individual subunits. The major difference between the two structures is a rotation of ~5–10° in the orientation of the exonuclease domain with respect to the thumb subdomain.

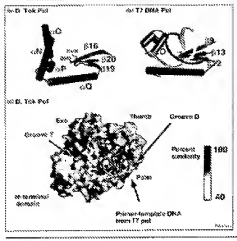
The domains of D. Tok Pol are arranged as an irregularly shaped flattened ring with a central cavity formed near the polymerase active site. The mostly α-helical thumb subdomain forms one side of the active-site cleft and makes contacts with the exonuclease domain (Figure 1). The structures of the thumb domains of various polymerases are often unrelated in sequence. However, in all cases where structures are available the thumb domain is seen to exhibit an important role by flanking primer/template duplex DNA as it exits the polymerase active site [6]. The D. Tok Pol structure has been determined in the absence of DNA, and a portion of the thumb subdomain that is likely to contact DNA structures (65–103) is disordered. This is commonly observed for the corresponding regions of other polymerases in the absence of substrates [6,11–14]. In the DNA polymerases from bacteriophage T4 and RB69, the thumb subdomains also provide a C-terminal sliding-clamp that interacts with the processivity factor [15,16]. In D. Tok Pol, the corresponding region (residues 757–773) is disordered.

The second region of the active-site cleft is occupied by the gene subdomain and includes residues important for substrate discrimination and the catalysis of the polymerase reaction. In D. Tok Pol, the palm is organized around seven β strands (β1a, β1b, β2b) flanked by six α helix (αD–αJ) (Figures 1 & 3a). It contains two disulfide bonds (Cys128–Cys142, Cys305–Cys359) that have not been previously observed in palm subdomains and which may be important for thermostability (Figure 3).

Figure 1

Structure of D. Tok Pol. The structure is represented by polymerase palm, exonuclease, and other motifs as ribbon backbone and secondary structure elements. Two proteomes, represent metal ions (produced by the Mg2+ coordinating the divalent to the exonuclease domain). The active-site residues Asp323, Asp325, Asp327, Asp405, Asp407, Asp408, and Asp410 were mutated. The two dimethylamines are represented. Regions of the polymerase chain that could not be modeled in the present structure because of disorder are indicated by dotted lines. The various domains and subdomains and their boundaries are indicated in the key.

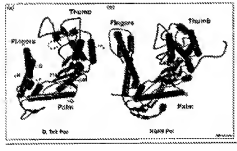
Figure 2



Comparison of DNA polymerase structures. (a) A view of the tetrameric structural elements of the polymerase which are shown (palm and fingers subdomains) of B. Tok Pol (also see Figure 1). (b) The corresponding regions of T7 DNA polymerase including the primer-template duplex from its ternary structure (PDB code: 1T8P) [21]. The orientation of T7 DNA pol was chosen to maximize the similarity between αP, αQ, and αD of B. Tok Pol. The view differs by about 90° from an analogous picture looking at the active site positions on T7 DNA pol. (c) A GRASP surface representation of B. Tok Pol, also oriented as the complex shown in the T7 DNA polymerase–DNA complex (PDB code: 1T8P) [21]. The surface is colored according to sequence identity [55–57]. Calculation of, as in Figure 3c. The primer strand is an orange, worm-like tube and the template strand is in gray.

The unusual structure of the palm subdomain from polymerases belonging to the Pol I and Pol II families can be aligned closely (the exit helix of the palm domains differ significantly) [6,7]. The conserved structural elements (β16, β19, β20 and helix αQ) are in the range of 0.6–2.0 Å, indicating a conserved conformation of functions. These are not restricted to the palm domains

Figure 3



Comparison of the structures of B. Tok Pol and RB69 Pol. The structures of B. Tok Pol and RB69 Pol are presented in ribbon representation to highlight the similarities and differences. Structural elements that are in common between the structures are represented in the same color as in Figure 1. Elements that are unique to RB69 Pol are colored magenta. Disordered segments are indicated by dashed lines. The N-terminal and the exonuclease domains are not shown.

of Pol I polymerases that are essential for polymerase activity, because they coordinate two metal ions [2,19,21,27]. The corresponding residues in D. Tok Pol are Asp406 and Asp543 (Figure 2). No metal ions are, however, evident in our electron-density maps.

The fingers subdomain in D. Tok Pol consists of a set of antiparallel α helices (αD, αE, αF, Figure 2). These helices are shorter in length than the corresponding elements of RB69 Pol and a helical segment that connects helices G and N in RB69 Pol is missing altogether (Figure 3). The fingers domain of D. Tok Pol is anchored to external structures to that of Pol I type polymerases (Figure 2). However, helix αF in D. Tok Pol is positioned distantly to helix O in Pol I polymerases (Figure 3), and is likely to play an analogous and essential role in recognition of the incoming nucleotide [9–11,28].

The PolB exonuclease domain in D. Tok Pol is located opposite the thumb subdomain and above the fingers subdomain, as noted for RB69 Pol. It contains two metal ions – presumably Mg^{2+} ligated to Asp411 and Glu141 (Figure 3). The position of this domain relative to the polymerase active site is distinct from the arrangement in a Pol I type polymerase. The conservation between RB69 and D. Tok Pol of the location of the exonuclease domain suggests that this is a characteristic feature of the B type polymerases. The presence of the D. Tok Pol 3′–5′ exonucleolytic domain resembles those associated with other DNA polymerases [29,30]. The 3′–5′ exonuclease domains from the Pol I (*E. coli*, *T. aquaticus*, *Bacillus subtilis*, bacteriophage T7) or Pol II (RB69) polymerase families can be aligned with each other closely (small to medium-size elements [31], [32], [34] and helices αB and αI in the range of 1.0–2.8 Å). This alignment suggests principles underlying associated with substrate binding, catalysis and metal binding in a exonuclease mutator (Figure 4) [6].

The arrangement of the N-terminal, exonuclease, and polymerase domains creates two deep grooves leading into and out of the polymerase active site. The D groove (for duplex-DNA binding, following the nomenclature of [7]) is a tunnel exclusively below the thumb subdomain and encircles a region of positive electrostatic potential. The T groove (for template/DNA binding) lead away from the active site in the opposite direction and is located below the fingers subdomain. A small β-sheet tube relating elements from the polymerase domain to the exonuclease wave was (Figure 2c).

We have used the structure of T7 Pol bound at primer-template DNA to model DNA into D. Tok Pol (Figure 2c). Superposition of the palm subdomains of the two polymerases shows that remarkably few but contacts are formed between the DNA from T7 Pol and amino in the D. Tok Pol model. The two regions that does collide

Structure alignment of exonuclease domains. Exonuclease domains from RB69 Pol (blue), D. Tok Pol (green), and T7 DNA polymerase (yellow). Helices αB (residues 197–213), αC (residues 237–243), αD (residues 267–283), αE (residues 287–298), and αH (residues 307–313) which participate in metal ion binding are shown in red. αA, a short helix that is conserved in various duplex-DNA exonucleases, is the helix of superimposes with only residues from helix αB (residues 203–213) is visible in D. Tok Pol. Structural comparisons can highlight conserved sequences and motifs which in catalysis are essential to proper DNA and RNA exonuclease. Two motifs, typically associated with metal ions, appear in the active site. The motifs are in close association to a motif which in its unliganded state has a configuration of the exonuclease domain from the RB69 Pol structure.

with the DNA in the minimum interactions the exonuclease and polymerase domains. The region (residues 372–388) is partially disordered in the T7 Pol structures, and is likely to recognize upon binding DNA. This superposition allows five base pairs of DNA to be accommodated in the D. Tok Pol active site, with the formation of DNA-protein contacts. The formation of contacts with additional base pairs would require a change in the position of the thumb subdomain to the region of the D groove. A change in the conformation of the fingers subdomain (between αG and αH) is also required to position residues Lys484 and Tyr489 (or Tyr490 of D. Tok Pol (Figure 2) for interaction with the incoming nucleotide, by analogy with the T7 Pol structure [11]. Finally, the super-imposed primer-template DNA is well positioned so that the incoming template strand will preferably reside in the T groove. Superposition of the DNA molecule derived from the structure of HIV-1 reverse transcriptase complexed to DNA [33] leads to similar conclusions.

Comparison between D. Tok Pol and RB69 Pol

Although the DNA polymerases from D. Tok Pol and bacteriophage RB69 share less than 30% protein sequence identity (Figure 5), their structures resemble each other

Figure 8

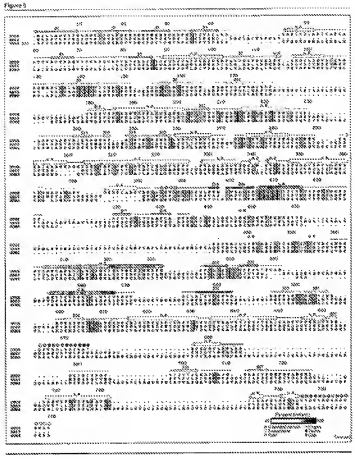


Figure 5

Structure-based sequence alignment of D. Tok Pol (DP1DM), RB69 Pol (RB69P) and human Pol δ (HUMPD). The highly sequence-homologous residue 372, as indicated by the number on the beginning of the sequence, the alignment is extended to sequence similarity (60% when no 100%, gaps calculated as described in Figure 3). Shown here is a small subset of a larger set of sequences that were used to generate the alignment. The full sequence alignment is available as supplementary material published with this paper. The sequences representing structural elements are indicated as in Figure 3 and accompanied by human numbering scheme for amino acid sequences. Gaps in the sequences represent portions of the polypeptide chain that could not be traced.

closely (Figure 3). Not surprisingly, the regions of highest structural similarity are concentrated in and around the exonuclease and polymerase active sites (Figures 2c,3). Despite the low overall sequence identity, the individual subdomains in the two structures superimpose well (rms error in Cα positions in the fingers, thumb and palm subdomains is in the range of 0.8 to 1.3 Å). Moreover, the overall arrangement of domains and subdomains with respect to each other is preserved in the two polymerases, strengthening the proposal that Pol II DNA polymerases share a common architectural plan (Figure 3).

One difference between the overall structures of D. Tok Pol and RB69 Pol concerns the orientation of the exonuclease domain with respect to the rest of the structure. When the two polymerases are superimposed via their respective palm subdomains it is seen that the exonuclease domain of RB69 is rotated inwards by ~8°, bringing the active site to a subtly asymmetric configuration [29]. In contrast, the exonuclease domain in D. Tok Pol (to 82.8Cα of the exonuclease engraved in ruins). It is possible that conformational change between open and closed configurations of the exonuclease domain are a part of the biochemical cycle

Figure 6

Comparison of surface charges in D. Tok Pol and RB69 Pol. Accessible surface representations of D. Tok Pol and (b) RB69 Pol. The two same molecules are in the orientation of the polymerase active site cleft in the front. Surface regions of positive and negative electrostatic potential are represented as contoured blue and red, respectively, contoured at ±10 kT/e. Surface regions corresponding to neutral potential are colored white. The figure was prepared using GRASP [30]. A representation of D. Tok Pol as a ribbon is also shown for orientation.

of the protein, particularly at the two different forms of D. Tok Pol (RB69 in the conformation of the exonuclease domain that binds).

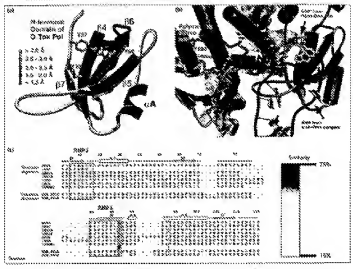
One interesting difference between D. Tok Pol and RB69 Pol is that the former is a thermostable DNA polymerase whereas the latter is not. Thermostability is thought to usually features in the D. Tok Pol structure that might be correlated with thermostability is complicated by the very low sequence similarity between the two enzymes. Even features that differ should not, however, in their tolerance limitations of amino acid interactions on the surface of D. Tok Pol when compared to that of RB69 Pol (Figure 6). The functions of networks of ionic interactions has been found to correlate with thermostability in other proteins [31,32].

Generally, D. Tok Pol polypeptide tend to be more compact with smaller helices and shorter loops than are found in RB69 Pol a feature that may be another important source of thermostability. For example, the palm subdomain displays close structural conservation of elements near the catalytic reaction residues. However, helix αR in D. Tok Pol is much shorter than in correspondent in RB69 Pol, and a small subdomain in front of the palm subdomain is entirely missing in D. Tok Pol (Figures 3,5). Deletion of these elements is also seen in exonuclease-less of archaebacterial DNA polymerase [33,34]. Likewise, the fingers subdomain is missing a large mass of loops as up in D. Tok Pol (Figure 3,5). However, the RB69 fingers extensive exon probably plays a TM phage specific role, as it is also missing from the sequences of archaebacterial DNA polymerases and eukaryotic polymerases (Figure 5).

The N-terminal domain resembles RNA-binding domains

The N-terminal domain of D. Tok Pol has no corresponding element in Pol I type polymerases. Analysis of the

Figure 2



The N-terminal domains of O Tox Pol. (a) Ribbon representation of RNA-binding domains, consisting of RNA-binding regions from R1-R5 (domains A and B). TPP1, TPP2, and TPP3 (brown) in R2 have been aligned by superimposing ALPHARNA, CUPRIPROT2.0. The Pol enzymes' R2–TPP1 region represent their variants. The R4, R5, R6–R8, and two helices (light gray) with the N-terminal domain of E. Tox Pol is shown. A cluster (yellow) of seven acidic residues in the R4 proteins at the vicinity of asymmetrical R4 molecule, ranging from Thr 1 to 5 Å to which Arg 48 (yellow) domain, particularly the Tox Pol helices are shown in red. (b) Stereoview of the R1–R5 domains of O Tox Pol. The R1 domain is colored in red from Met 1 to Ser 127 from Met to Ala 1. But the RNA core was from Met 1–R4–R5 (orange) RNA core TRPR. RNA molecule with the N-terminal domain of O. Tox Pol. The model was generated by superimposing the DNA RNBP-1 and the N-terminal domain of O. Tox Pol using Cα backbone atoms (yellow). The RNA is shown in blue with the major and minor backbone representations as a worm and the bases as wall and stick representations. A partial surface that represents the interface between the N-terminal domain of E. Tox Pol and the nucleotide binding is shown in gray. The location of the mutated RNA residue in the polymerase active site is displayed by marking the position of cross-Trp54. The sequence alignment also represented as the glutamate homeodomain (GLMF) molecule found as the representative 6800 of RNBP-1. overall an right groups mostly domains with the positions of one binary of the negative RNA main loop. (c) Sequence and primary sequence alignments of RNA-binding domains. Sequence alignments of the R1 N-terminal domains from E. Tox Pol and RNBP-1 (fluorescence domain) and the RNBP-1 from Met 1 (domains A and B). TPP1, TPP2, TPP3 (domains A, B, and C) from Met 1 to Ser 127. Alignments of the N-terminal domains of E. Tox Pol and the RNBP-1 sequence identity were obtained using CLUSTALX (EMBL, using the default parameters). The conserved primary sequences from 1–50 and 60–90 are boxed. The alignment is colored by sequence similarity (75%, white to 15%, green) calculated by averaging the similarity scores at each position in all possible pairs of sequences (the unconditional scheme 4). Equivalence of the sequence residues are established by use of the RNBP-1 and E. Tox Pol sequences. Numbering of residues and naming of secondary structural elements is that of O. Tox Pol.

interaction of this domain using [3H]-[32P] [35S] chemical-crosslinking specificity revealed a previously unrecognized similarity of RBDs. RBDs are small nucleotide-

(80–90 residues) found in RNA-binding proteins of prokaryotes, archaea, and eukaryotes. Reviewed in [18]. These modules adopt a conserved four-stranded antiparallel

and bind to single-stranded RNA. Two conserved sequence motifs, referred to as RNP1 (phenylalanine-rich) and RNP2, provide aromatic and charged residues that are responsible for RNA recognition [35,36] (Figure 7).

The N-terminal domain of D. Tok Pol can be superimposed closely over the core secondary structural elements of RRMs from the U4A spliceosomal protein [35], ribosomal protein S6 [36], the heterogeneous ribonucleoprotein (hnRNP) protein from HIV-2 (human) [37,38] and the aminoacyl-tRNA synthetase from *E. shimogushiki* phenylalanyl-tRNA synthetase [39]. The results in Cα positions for these superpositions are in the range of 4.5–2.9 Å (Figure 7a). Differences between the structures of the loops in the N-terminal domain of D. Tok Pol and those of the RNA-binding domains are within the range of structural variation seen in the various RNA-binding domains.

There is no evidence at present to suggest that the N-terminal domain of D. Tok Pol binds RNA. However, comparison with the structures of RNA complexed of RNA-binding domains shows that the N-terminal domain might in fact be a functional RNA-binding domain (Figure 7). In particular, three aromatic residues in the N-terminal domain (Phe37, Tyr59 and Tyr80) could interact with RNA bases in a manner similar to that seen in crystal structures of RNA bound to RNA-binding domains [35] (Figure 7). Interestingly, these residues are located near the junction of a guanosine triphosphate molecule and a flavin-bound metal N-terminal domain of RBP9 Pol II (Figure 7b). The DNA polymerases from bacteriophage T4 and its distant relative bacteriophage RB69 bind specifically to the ribozyme-binding site of their own mRNA (messenger RNA), repressing its translation [40–42]. The N-terminal domains of T4 Pol and RB69 Pol are smaller than that of D. Tok Pol. In the RB69 Pol structure, the N-terminal domain seems to form an incomplete RNA-binding domain (Figure 7c).

There is no significant overall sequence similarity between the N-terminal domain of D. Tok Pol and RNA-binding domains, which is why the presence of this fold was not recognized previously (Figure 7a). Comparisons of the sequences of other archaeal/viral DNA polymerases and human polymerases δ and ε suggest that a similar three-dimensional structure is likely to be found in these polymerases as well (Figure 7a). The sequence alignments in this region is unambiguous for the archaeal/viral DNA polymerases. For eukaryotic polymerases the alignments is less certain, but it is easier to conserve the conserved aromatic character of the RNP motifs (Figure 7a). Confirmation of the presence of these domains along with more ability to bind RNA and their precise role in eukaryotic DNA synthesis awaits future structural and functional studies.

Biological implications

The structure of the DNA polymerase from the archaebacterium *Thermococcus* strain Tok, D. Tok Pol, reveals a strong similarity to the DNA polymerase from bacteriophage RB69. It also reveals the presence of an N-terminal domain that has structural similarity to RNA-binding domains from the U1A spliceosomal protein, ribosomal protein S6, the hnRNP protein and the aminoacyl-tRNA synthetase. Although the structure is the accessible vicinity of the central catalytic region of the polymerase domain closely resembles that of Pol I type DNA polymerases, the overall architecture of D. Tok Pol and the placement of the exonuclease domain is strikingly different. The similarity between D. Tok Pol and RB69 Pol suggests that these two structures are representative of a common Pol II polymerase fold. Members of this family carry out chromosomal DNA replication in archaeans, including humans, and yet there is no structural information available for any eukaryotic member of this family. While this manuscript was being prepared, the structure of another archaeal/viral DNA polymerase, that from the organism *Thermococcus gorgonarius* has been reported [43]. The D. Tok Pol structure reported here, along with the RB69 Pol structure and the structure of the *Thermococcus gorgonarius* DNA polymerase, should now make it possible to generate reliable structural models for eukaryotic DNA polymerases.

Materials and methods

Protein expression and purification

The D. Tok Pol bacterial expression vector and initial strain and sequence were generous gifts of Dr. Hyeoung-Chyn Kim. Cultures are reproducible. Lysate expression was achieved by the co-expression of the gene encoding the DNA polymerase through the use of the arabinose repressor of the polymerase was completed using the plasmid pBAD/Myc-His A (Invitrogen). Cells were grown in the above expression system in a 10 liter, pan-type cell fermentor. The cells were harvested by centrifugation at 5000 rpm for 30 min and further purified by the following steps. The first step purification step was heat denaturation at 75 degrees by incubation. Subsequent to 70 mM $NaCl$, pH = 7.5], 50 mM $Tris_2HCl$, for crystallization from.

Crystallization, characterization, and heavy-atom derivatization

Crystals of D. Tok Pol (protein concentration, 10 mg ml^{-1} 100 mM $NaCl$, were prepared from 100 mM $TRIS$-HCl (pH = 8.0), 100 mM $MgCl_2$, 200 mM $(NH_4)_2SO_4$, 30% (w/v) 2-methyl pentane-2,4-diol (MPD). The best crystals were obtained in vapor diffusion at 18°C. These crystals were crystallized in 100 mM $TRIS$-HCl (pH = 8.0), 10 mM $MgCl_2$, 200 mM $(NH_4)_2SO_4$, 30% (w/v) MPD, 10% (w/v) PEG400. Several other crystallization conditions were screened. Crystals were grown from hanging drops containing 2 μl of protein solution and 2 μl of reservoir solution equilibrated against 1 ml of reservoir solution at 18°C. The crystals belong to space group $P2_12_12_1$ with unit cell dimensions $a = 87.0$ Å, $b = 107.0$ Å, $c = 120.0$ Å. Native crystals diffract to at least 2.9 Å resolution on a rotating anode x-ray source. There are two molecules per asymmetric unit, although crystal contact faces also exist that could accommodate a crystal with one molecule per asymmetric unit. Two heavy-atom derivatives were obtained by soaking native crystals in mother liquor containing 1 mM K_2PtCl_4 for 18 h. Data were collected at room temperature using a Rigaku R-AXIS IIc image plate area detector mounted on a Rigaku RU-200 rotating anode x-ray generator operating at 50 kV and 100 mA. Data were processed using DENZO and SCALEPACK [44]. Native data to 2.9 Å resolution were collected from a single crystal. Data statistics are given in Table 1.

RNP, RNase A, and DNase I, $a = 103.6$ Å, $b = 103.6$ Å, $c = 150.9$ Å, $\alpha = 90°$, $\beta = 90°$, $\gamma = 90°$. Three-wave datasets were obtained by scanning across a single crystal in the beam, collecting 90 MAR Research images, each 2° in oscillation.

Data collection and photon-counting histograms

X-ray diffraction data sets from a pair of electron-density maps and anomalous maps were collected as described in the Cornell High Energy Synchrotron Source (CHESS) beamline A1 [5, 6, 7]. Data from Native 1 crystals (tetragonal unit cell) were collected by a 'step' scanning of 2.4° with $\lambda = 1.000$ Å. 180 images were accumulated at 90 s per exposure with 512×512 pixels yellow zone by Mylar, 2.8 Å X-ray diffraction data, with average crystalline and solvent mosaicity each 0.6°, averaging 0.7°.

The processing of these data were scanned manually by inspection of diffraction images and checked by cross-checked difference Fourier maps. Experimental phases were calculated using these data with the program SHARP [8-9]. In our terms, native quality electron-density maps were obtained by performing anomalous single isomorphous replacement (SIR) calculations at 'SHARP' and examining the refined SIR phases with using the program SIGMAA [10,11]. The experimental phases were improved and extended by solvent flipping and negative-density truncation as implemented in SOLOMON [12]. The produced SHARP/SIGMAA/SOLOMON phased electron-density maps of sufficient quality to allow for initial Cα tracing of polypeptide chains automatically. This step was dramatically improved over a map interpretation with ref-manual/RIBBONS/O/CHAIN [13].

Model building and refinement

The initial model was built by hand with a 3.0 Å electron-density map using the interactive molecular graphics program O [14]. Model refinement was carried out using simulated annealing torsion parameter dynamics and restrained positional and temperature factor refinement with X-PLOR [15]. Refinements against the 2.8 Å Native 1 data set was carried out using data extracted at the inside aspects E_0 extended electron-density maps using $|2F_o|-|F_c|$ and $|F_o|-|F_c|$ coefficients, calculated by omitting selected solvent or ion molecules. Each cycle of refinement was carried out using the program X-PLOR [15]. The progress of the refinement was monitored by inspection of R_{free} (10% of the observed reflections) [16]. During the Native 1 step set, the model was refined to an $R_{free} = 29.6\%$ and $R_{cryst} = 24.1\%$. The refinement was continued further for 2.4 Å data. Native 1 data set is significantly worse in CNS than the 2.8 Å native) refined a new model that was written as above. The final results for Native 1 was refined to an $R_{free} = 29.0\%$ and $R_{cryst} = 23.3\%$, and the final native protein, solvent, and ion model. The model contains 3503 protein atoms, 1 chloride ion, 1 sodium ion, 3 magnesium ions, and 116 water molecules. The Native crystal contains 3503 non-solvent protein atoms, 8 sulfate ions, 2 magnesium ions, and 486 water molecules. Model geometry was analyzed using the program PROCHECK [17]. Most residues have an excellent φ-ψ Ramachandran plot, with over 90% of the residues in the most-favored region.

Figure preparation

Figures were composed in programs PGPLOT [18] and GRASP [19]. Stereo pairs were generated with BOBSCRIPT [20]. Figures 3 and 5 were rendered using RASTER3D [21].

Accession numbers

Coordinates have been deposited with the Brookhaven Data Bank by the Structure Biochemistry, under the accession code 1CSE.

Acknowledgements

We thank Dr. Tahir Hussain, Ph. for assisting the cloning, partial sequence determination and initial protein preparation of Cl. fer. DNA sequence determination performed by the Protein/DNA Technology Center of the Rockefeller University. We thank S. Burley, for technical assistance; J. Berman, B. Sweet, and the staff of the Wallace Synchrotron Type Source and the Cornell High Energy Synchrotron Source (CHESS) for technical support for assistance and instruction with data collection; and Richard Kriwacki, Jeffrey Peterson, Jeffrey Pelletier, Nikola P. Pavletich, Nikolaos P. Pavletich, Stephen Burley, and Tahir for help and advice. This work was supported in part by a short term Research Initiative of Merck (SB/93/1/14). Work at the Cornell High Energy Synchrotron Source (CHESS) is supported by the National Science Foundation under award DMR-9311937 with active. Mac-CHESS-beamline access (MHATS) was supported from the Office of Health Affairs, Department of Defense, and a Burroughs Wellcome Fellowship.

References

1. Grzeskowiak, K. & Dickerson, R.E. (1989). Comparison, volume of, and hydration characteristics of DNA polymorphs. *Biochem. Biophys. Res. Commun.* **164**, 1-8.
2. Privé, G.M. & Sadler, J.A. (1994). Function and structure determinants in DNA polymerase. *Annu. Rev. Biochem.* **63**, 779-821.
3. Jones, O.W. & Kaiser, I.H. (1966). Preparation, structure and activity of polynucleotides of known composition. *J. Biol. Chem.* **241**, 1537-1543.
4. Shamoo, C.R. & Steitz, T.A. (1999). Structure and enzymatic properties of a complex between the Klenow fragment and DNA: one nucleotide remaining. *Cold Spring Harb. Symp. Quant. Biol.* **64**, 63-71.
5. Tolan, J. & Fisher, T. (1999). Solving crystal structures and their function components. *Curr. Opin. Struct. Biol.* **9**, 21-28.
6. Arnold, E., Ding, J., Hughes, S. & Hostomsky, Z. (1999). Structure of HIV-1 RT/DNA complex and implications for drug resistance. *Curr. Opin. Struct. Biol.* **9**, 27-35.
7. Steitz, T.A., Ollis, D.L., Hendrickson, W.A. & Konigsberg, W.H. & Steitz, T.A. (1982). Crystal structure of a DNA-dependent RNA polymerase from bacteriophage RB69. *Cell* **69**, 1035-1041.
8. Bonner, J. & O'Donovan, M. (1994). DNA-protein interactions and enzymology. *Curr. Opin. Genet. Dev.* **4**, 765-769.
9. Qiu, S., Shi, A., Pavletich, P. Steitz, N.P. & Steitz, T.A. (1994). Structure of the Klenow fragment of DNA polymerase I complexed with deoxynucleoside triphosphate and pyrophosphate. *J. Mol. Biol.* **242**, 736-746.
10. Li, Y., Korolev, S. & Waksman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* **17**, 7514-7525.
11. Doublié, S., Tabor, S., Long, A.M., Richardson, C.C. & Ellenberger, T. (1998). Crystal structure of a bacteriophage replicative-type DNA polymerase. *Nature* **391**, 251-258.
12. Golden, B.L., Gottesfeld, J.M. & Hollenbeck, J. (1999). Architectural DNA polymerase specific transcription factor IIIC. *J. Biol. Chem.* **277**, 31080-31088.
13. Steitz, T., Smerdon, S., Jan, M. Kuriyan, P. & Kelly, J. (1985). Structural implications for the catalytic mechanism of the bacteriophage polymerase from the protease. *Proceedings National Academy Sciences, Acad. Sci. USA* **82**, 8978-8982.
14. Kelly, T.J. Hansen, R.D. & Lee, W.R. (1981). Characterization of a DNA polymerase from the nucleoprotein-associated Drosophila melanogaster. *Proc. Natl. Acad. Sci. USA* **78**, 6696-6700.
15. Brünger, A.T. (1992). X-PLOR. A System for X-ray Crystallography and NMR. *Yale University Press, New Haven, CT*.
16. Brünger, A.T. (1992). Free R value: cross-validation in crystallography. *Methods Enzymol.* **277**, 366-396.
17. Laskowski, R., MacArthur, M., Moss, D. & Thornton, J. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283-291.
18. Pearson, W.R. & Lipman, D.J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85**, 2444-2448.
19. Nicholls, A., Sharp, K.A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281-296.
20. Esnouf, R.M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* **15**, 132-134.
21. Merritt, E.A. & Murphy, M.E.P. (1994). Raster3D Version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 869-873.

11 Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340: 783–795

12 Bendtsen JD, Kiemer L, Fausbøll A, Brunak S (2005) Non-classical protein secretion in bacteria. BMC Microbiol 5: 58

13 Bendtsen JD, Nielsen H, Widdick D, Palmer T, Brunak S (2005) Prediction of twin-arginine signal peptides. BMC Bioinformatics 6: 167

14 Bernstein HD, Poritz MA, Strub K, Hoben PJ, Brenner S, Walter P (1989) Model for signal sequence recognition from amino-acid sequence of 54K subunit of signal recognition particle. Nature 340: 482–486

15 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2003) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 185: 4105–4116

16 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

17 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

18 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

19 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

20 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

21 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

22 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

23 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

24 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

25 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

26 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

27 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

28 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

29 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

30 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

31 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

32 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

33 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

34 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

35 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

36 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

37 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

38 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

39 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

40 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

41 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

42 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

43 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

44 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

45 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

46 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

47 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

48 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

49 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

50 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

51 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

52 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

53 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

54 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

55 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

56 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

57 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

58 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

59 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

60 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

61 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

62 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

63 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

64 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

65 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

66 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

67 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

68 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

69 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

70 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

71 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

72 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

73 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

74 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

75 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

76 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

77 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

78 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

79 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

80 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

81 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

82 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

83 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

84 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

85 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

86 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

87 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

88 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

89 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

90 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

91 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

92 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

93 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

94 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

95 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

96 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

97 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

98 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

99 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

100 Bhasin N, Messner P, Graninger M, Carlin A, Schäffer C (2004) Identification of a gene essential for O-acetylation of the S-layer glycoprotein in Geobacillus stearothermophilus NRS 2004/3a and preliminary characterization of the modified glycan backbone. J Bacteriol 186: 2672–2682

Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 3669–3674, March 1999
Biochemistry

Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*

Key words: archaea/DNA double-branch/replication/mechanism/exonuclease

KARL-PETER HOPFNER*†‡, ANDREAS EICHINGER*, RICHARD A. ENGH*§, FRANK LAUE¶, WOLFRAM ANKENBAUER¶, ROBERT HUBER*, AND BERNHARD ANGERER¶||

*Abteilung Strukturforschung, Max Planck-Institut für Biochemie, D-82152 Martinsried, Germany; and ¶Roche Diagnostics, D-82372 Penzberg, Germany

Contributed by Robert Huber, January 21, 1999

ABSTRACT Many known archaeal DNA polymerases belong to the type B family, which also includes the DNA replication polymerases of eukaryotes, but molecular high fidelity in extreme conditions. We describe here the 2.5 Å resolution crystal structure of a DNA polymerase from the archaeon *Thermococcus gorgonarius* and identify structural features of the fold and the active site that are likely responsible for its thermostable function. Comparison with the mesophilic B type DNA polymerase gp43 of the bacteriophage RB69 highlights thermophilic adaptations, which include the presence of two disulfide bonds and an enhanced electrostatic complementarity at the DNA-protein interface. In contrast to gp43, several loops at the exonuclease and thumb domains are more closely packed, this apparently blocks primer binding to the exonuclease active site. A physiological role of this "closed" conformation is unknown but may represent a polymerase mode, in contrast to an editing mode with an open exonuclease site. This archaeal B DNA polymerase structure provides a starting point for structure-based design of polymerases in fundamental applications in biotechnology and the development of antiviral or anticancer agents.

Propagation of cells requires faithful DNA replication. This is performed in vivo by DNA polymerases (pols), which select the appropriate dNTP at the nascent DNA primer strand to match its paired template. Different families of pols are encoded in different DNA polymerization processes including not only DNA replication (1, 2) but also repair and recombination (3, 4). A heterogeneity also manifested by varying polypeptide structures and/or subunit compositions (5, 6). Some pols strengthen polymerase activity with $3' \rightarrow 5'$ exonuclease activity (proofreading activity) within $5' \rightarrow 3'$ "exonuclease-specific endonuclease" activity often located in extensive structural elements on the same polypeptide chain (6–8).

Crystal structures are available for seven known polymerase families, including the A family DNA polymerases (9–14), pol β (15–17), HIV reverse transcriptase (18–20), and recently, the B family pol gp43 from bacteriophage RB69 (21). All show a functional polymerase structure, which resembles a right hand built by the palm, fingers and thumb domains (see ref. 7 for review). Although the fingers and thumb domains are highly divergent among the different families, the palm domains, which contain the conserved catalytic/aspartate residues, show a similar topology among all families, except pol β. The polymerase nucleotide transfer site studied in detail for the A family polymerases, HIV reverse transcriptase, and pol β, and was shown to involve two metal ions (summarized in ref. 7).

Considerably less is known for the family of type B pols, which are replicative enzymes in eukaryotes and some thermoacidophilic archaea (22, 23). The structure of gp43 from bacteriophage RB69 (21) provided an excellent first insight into this family in addition to the three polymerase domains, gp43 contains an $3' \rightarrow 5'$ exonuclease domain and an N-terminal domain. The exonuclease and palm domains share the topology and active site of A family enzymes, implying similar metal-assisted mechanisms for polymerase and exonuclease activities (21). The thumb and finger domains are apparently specialized to the other polymerase families. The function of the N-terminal domain remains unknown, but may help to confer the unique replisome replication apparatus (21).

Much is known about the replication of phages [24–26], viruses [1, 27], Protozoa, [28] and Eubacteria [1, 3, 29, 30], which in general involves pols that also process primers, RNase-H, sliding clamps, and other factors (31). Considerably less is known for archaeal replication, where mostly B type polymerases, similar to eukaryotic replication enzymes pol α and δ, have been identified (6, 22, 23, 32–34). The relative abundance is surprising, because such crucial biotechnological applications as cloning and PCR require the thermostability and fidelity typical of archaeal polymerases (35). Thus, in addition to satisfying basic research interests, structural information could aided, for example, the engineering of archaeal enzymes with tailored nucleotide incorporation rules or the design of antiviral and anticancer polymerase inhibitors. For these reasons, we have determined the structure of a DNA polymerase from *Thermococcus gorgonarius* (Tgo), an extremely thermophilic archaeal archaebacteria isolated from a geothermal vent in New Zealand (36). The enzyme possesses pol and $3' \rightarrow 5'$ exonuclease activity, which together ensure thermostable replication with high fidelity (error rate: 3.5×10^{-6} to 2.3×10^{-7}; see ref. 36). The 2.5 Å structure shows a topological similarity to gp43 and gives insight in the structural biology of archaeal DNA polymerases, including the identification of special mechanisms for thermostable adaptation.

MATERIALS AND METHODS

Materials. All materials were of the highest grade commercially available.

Bacterial Strains. *Escherichia coli* BL21(DE3) containing pUBS520 was used as described (36). *E. coli* JM84 (DE3) (this study) was a generous gift of Matthias Budisa (Max-Planck-Institut).

Expression Vectors. PSTag1 was obtained from Roche Molecular Biochemicals.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

Abbreviations: Tgo, *Thermococcus gorgonarius*; DNA pol, DNA polymerase; Dam, deoxyadenine. The atomic coordinates have been deposited in the Protein Data Bank, Brookhaven National Laboratory (Upton, NY) (file name 1TGO).

†Present address: Gene Center, Ludwig-Maximilians-Universität, Feodor-Lynen-Strasse 25, 81377 München, Germany.

‡To whom reprint requests should be addressed. e-mail: hopfner@genzentrum.lmu.de.

§Present address: Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, 6-155 Jackson Hall, 321 Church Street SE, Minneapolis, MN 55455.

||Present address: Roche Molecular Biochemicals, Nonnenwald 2, 82377 Penzberg, Germany.

Table 1. Data collection and nonreundance refinement

Derivative	Resolution limit, Å	Total observations	Unique observations	Completeness, %	R_{sym}, %	R_{iso}, %	Phasing power
Native	3.0	136,915	21,379	91.9	7.9		
UO	3.2	75,964	20,119	99.5	11.0	18.1	0.71
PT1	3.2	65,836	17,955	96.2	12.0	27.5	1.34
PT2	3.0	53,870	6,910	68.9	14.2	12.8	1.17
PT3	3.4	90,336	14,105	92.8	10.5	17.1	1.16
PT4	3.5	107,827	25,089	99.6	7.7	18.5	0.67
PT5	2.3	328,690	26,585	88.7	6.1	36.2	0.83
PT10	3.5	39,629	11,457	93.4	9.6	38.7	1.24
PB	3.5	78,186	13,206	97.9	11.7	17.5	0.43
PMT	3.4	94,532	14,155	99.6	8.9	19.6	1.04
OS	2.8	188,168	24,036	91.2	5.2	28.4	1.81

Outermost layers of each (3.0–3.0 Å; 3.0–3.2 Å).

Heavy-atom derivatives were prepared by soaking the crystals in low salt buffers containing the heavy atoms in solutions 0.5–0.8 mM except for 2 h, PT1, 3 mM K$_2$PtCl$_4$ for 1 h; PT2, 3 mM K$_2$PtCl$_4$ for 2 d; PT3, saturated room-temperature solution of EtHgCl for 1 d; PT4, saturated room-temperature solution of EtHgCl for 1 d; PT5, 3 mM K$_2$PtCl$_4$ for 3 d; PT10, 3 mM K$_2$PtCl$_4$ for 1 d; PB, 3 mM K$_2$PtCl$_4$ for 1 d; PMT, PT + PT5; OS, saturated EtHgCl for 1 d. PT1, PT2, PT3, PT4, PT5, PT10, PB and PMT were collected with a MAR imaging plate and the data were processed with DENZO (Otwinowski, Z. & Minor, W. (1997) Methods Enzymol. 276, 307–326). Native and OS were collected with a Mar345 image plate and the data were processed with DENZO. Hamburg.

Cloning, Expression, and Protein Purification. The gene for the MSB-like DNA-dependent pol (Tpo pol) was cloned from *Thermoplasma acidophilum* (ATCC 25905) genomic DNA and expressed in *E. coli* LE392pUBS520 plus pTpo (Expression Systems, Davis, CA). Tpo pol was purified essentially as described (30). The cell suspension in the TSE lysis buffer (Expression Systems) was sonicated on ice and with an additional concentration step on Porin 50 HQ anion exchange medium (Bayer, Molecular Biochemicals). Active fractions were combined, concentrated to 20 mg/ml, and transferred to 20 mM sodium phosphate, pH 8.2/10 mM 2-mercaptoethanol/0.02 mM MnCl$_2$.

The gene for a subunit/dimerization-containing subunit of Tpo pol (the Tpo pol) was expressed in *E. coli* BL21 (DE3) (Novagen) using established protocols (17). The Tpo pol was purified by using the same protocol.

Crystallization. Crystals of purified Tpo pol (or Tpo pol) were grown by using sitting drop vapor-diffusion techniques at 4°C with high-salt conditions [2.6 M protein/reservoir—100 mM Tris, pH 8.2/2.0 M ammonium sulfate] and cultivated to 3.0 Å (freezing) and to 2.8 Å (freezing; 30% vol/vol Ethylene Glycol/saturated Ammonium Sulfate/pH 8.2, Hampton Research, Laguna Niguel, CA). Native crystals (300 mM Tris, pH 7.8/2.0 M ammonium sulfate/20% PEG 400) yielded only poorly diffracting crystals but allowed soaks (including heavy atoms) with some cell constants accordance to a, b, c = 65.0, 105.0, 196.3 but unidentical sets of reflections.

Data Collection and Processing. Data were collected with a MAR imaging plate or a Bruker AXS X2000 mounted on a Rigaku rotating anode source, or with a MAR imaging plate on a MAR CCD (charge-coupled device) at beamline BW6 at DESY, Hamburg. The data were processed with DENZO (Molecular Applications, Inc., Glendale, CA) or SCALEPACK (MAR imaging plate) or HKL, scaled with SCALA (MAR CCD), and truncated with TRUNCATE (49).

Structure Determination. The structure was solved by multiple isomorphous replacement and anomalous scattering (MIRAS) by using data from crystals transformed to low salt conditions (Table 1). Cross-dispersion calculations were done with programs from the CCP4 suite (46). Heavy-atom positions of major sites were located in difference Patterson maps and were refined and improved (MLPHARE) to calculate protein phase angles to 3.5 Å resolution. A partial polyalanine model was built into noncrystallographic averaged electron density maps of the solvent mask by using MAIN (43). The quality of the electron density was improved by phase combination of the partial model with the experimental phases by using SIGMAA (46), and several cycles of solvent flattening to 3.0 Å by using DM/RESOLVE (46). At this stage, no noncrystallographic symmetry was found for a significant portion of the molecule, corresponding to about 167–234, 283–306, 467–728, and 753–773.

Model Building and Refinement. The partial model (R factor 29%) was tried to phase the 2.3 Å resolution data of the Tpo pol (high salt conditions). The model was oriented with another (46). The rotation/translation coefficient of 21.6% and the R factor of 58.3% showed divergence of the high- and low-salt structures. After both subunits (monomer, monomer B factor correction and rigid-body minimization) forming five dimers independently, the partial model was iteratively refined and extended with simulated annealing, Powell minimization, restrained individual B factor refinement with CNS (42), and manual model building with MAIN (43) by using data from 25.0–2.5 Å resolution (Table 2, Fig. 1).

RESULTS AND DISCUSSION

Structure of Tpo pol. Tpo pol is a long chained molecule with dimensions 80 Å × 90 Å × 190 Å. The single polypeptide chain of 773 aa is folded into two distinct structural domains (Fig. 2): the N-terminal domain (residues 1–280, the 3′ → 5′ exonuclease domain 131–378), the palm (369–499 and 540–582), fingers (500–539), and thumb (380–778) domains of the polymerase unit, and a helical interdomain insertion (281–368)

Table 2. Crystallographic refinement, Alphozet dimer

Space group	$P2_12_12_1$
Cell dimensions, Å	a = 80.1, b = 189.2, c = 194.2
Observations, 25–2.5 Å	
Total	482,446
Unique	79,421
Completeness, %	
Total	93.1
Last shell	86.6
R_{sym}, %	
Total	7.3
Last shell	20.2
R factor (R_{cryst}), %	20.9 (25.1)
rms deviation in bond lengths, Å	0.008
rms deviation in bond angles,°	1.5
No. of noncrystallographic axes	
Protein	6,378
Water	320

FIG. 1. Stereoscommonstrator of the electron-density map. The 3.0Å ~ F_o electron density (contoured at 1.0 σ) at the polymerase active site is well defined by the refined atomic model superimposed.

between the exonuclease and palm domains. The polymerase cleft lies on the DNA-binding crevice, reminiscent of a right hand, which is the overriding characteristic of pols. Opal 3 from bacteriophage RB69 also shows this overall domain topology (23).

These clefts extend radially from the polymerase active site at the center of the ring: two of them, in opposite directions, forming a major cleft across the molecule, and one perpendicular to those. Based on sequence homology to pol A family enzymes, the two opposite clefts probably bind duplex DNA (cleft D, according to ref. 23) and single-stranded template DNA (cleft T), respectively. The perpendicular (trifling) cleft binds the polymerase active site and the exonuclease active site and breaks the template strand in melting mode (11).

The exonuclease domain is structurally equivalent to the 3' → 5' exonuclease domain of pol A family (43). Like gp43, however, it is located at the opposite side of the polymerase core: by sequence comparis to the thumb domain at the melting cleft, on one side, and by a mass of and noncovalent contacts to the N-terminal and palm domains and the 42 residue interdomain helix, on the other side. This latter segment is located at the base of cleft T, which is additionally bounded by the exonuclease, 70 kDa domain, and palm subdomains.

The topology of the palm domain is conserved among polymerases Klenow (5), with two long helices (Q and R)

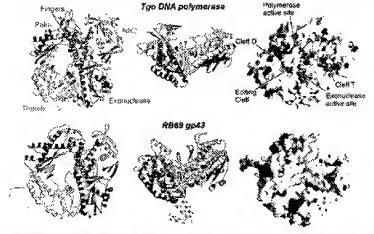


FIG. 2. Structure of Tgo pol and comparison with gp43 from bacteriophage RB69. Half-tone ribbon representation of Tgo DNA polymerase (figure at two orthogonal views with labelled secondary structure elements). The molecule is composed of five domains. N-terminal domain (yellow); 3' → 5' exonuclease (red); palm (light and dark magenta segments each associated with C-terminal segment, respectively (see text)); fingers (blue); and thumb (green), which are arranged in form a ring. An interdomain helical segment between the exonuclease domain and the palm is strange. The associated carboxy-terminal in the active site and the two specific cleaves are shown. The active site of Tgo is at the bottom, respectively. Tgo pol has the same overall architecture and domain topology than gp43 of RB69 (lower). The fingers, palm, exonuclease, and thumb domains of both enzymes are labeled as in Tgo. Comparison of molecular surfaces of Tgo pol (upper) and RB69 gp43 (lower). Red and blue indicate negative and positive electrostatic surface potentials, respectively. In contrast to gp43, Tgo pol has a strongly enhanced positive potential at the putative DNA-binding cleft.

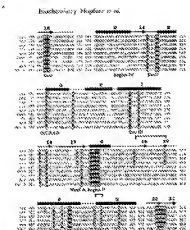


FIG. 3. Sequence alignment of B family DNA polymerases. The alignment was hand-adjusted from set II to highlight specific residues from the class of archaeal pols. The secondary structure of Tgo pol is indicated on top of the alignment with helices (black), sheets (hatched) and loops (thin) colored according to domains with the same color code as Fig. 2B. Strictly conserved residues of type B polymerases are red, and underlined residues are yellow. Universally conserved residues of archaeal type B enzymes—as discussed in the text—are green. Dnaf-like bands are shown by + for on top of the alignment. Abbreviations are: *Thermococcus gorgonarius* pol, Tgo; *Pyrococcus furiosus* pol, Pfu; *Thermococcus* sp. pol 4, Thermo4; *Methanococcus jannaschii* pol, Mja; *Methanobacterium thermoautotrophicum* pol, Mth; *Methanococcus vulnare* pol, Mvu; *Escherichia coli* pol II, pol II; *Escherichia coli* pol II, Ecopol II.

stacked against the five-stranded antiparallel β-sheet that provides the three conserved aspartate residues involved in nucleotidyl transfer. The fingers emerge from the palm domain as an α-helix-rich insertion. Set 90 residues are folded into two segments that cradle a cylinder of upper-helix (recall that, helix F contains the conserved Ka/Mo/Y/Dx(4) motif of B type polymerases) and is inferred to the O helix of A type enzymes (see below). The ~26-residue insertion between helix O and F in B/RB69 and T4 pol's is missing in Tgo pol, which folds helices and of smaller loops are much shorter than their equivalents in pol3. The shorter fingers of Tgo pol presumably reflect the typical structure of the archaeal/eukaryotic B type fingers (set in, and E, and B, and pol II). The thumb domain topology, similar to that of pol6, is correlated to other polymerase types. However, like the thumb of A type enzymes, a bundle of α-helices set in base proximal from the active site β-sheet floats in the screw like the thumb contains a 75-residue subdomain (565–729), which lines the solvent-faced domain and contributes to the exiting channel, explaining why mutations in the noncatalytic domain of B-type polymerases affect the polymerase activity and vice versa (ref. 45).

Weakly defined density across the cleft of the thumb domain was modeled as the C-terminal 6 residues with a polyalanine chain. The C terminus must then run proximal from the core exonuclease as in the RB69 polymerase (21). Because the C terminus of the T4 pol was involved in sliding-clamp binding (45), it is likely, however, that these residues become ordered on the same location as for Tgo pol.

Sequence Alignment of Archaeal DNA Polymerases. The structure of Tgo pol allows the generation of a structure based sequence alignment of the archaeal subfamily of type B DNA polymerases: the location of conserved and unique residues, and the comparisons with other type B DNA polymerases (Fig. 3).

Polymerase Active Site. The polymerase active site is formed by the central β-sheet extends 16, 17, 22, 11, and 22 aa (helix N of the palm domain and helix F located in the fingers and is highly conserved among B family polymerases Fig. 4). Three carboxylates required for nucleotidyl transfer in B family polymerases, two of which coordinate two metal ions (30) are exceptionally conserved among A family enzymes, B family enzymes, and reverse transcriptases (31). Superposition of Tgo pol and T7 replication complex (19) places the dNTP near the proposed nucleotide-binding site in helix F, Thr ^{540}Pro^{541}Thr^{542} motif (Fig. 5) and suggests interactions of the carboxylates with metals and the phosphoryl tail of the bound dNTP (Fig. 4). Superimposition of the metal-ion coordinated Tgo-587 aligns it to minus the Lys-522-phosphate tail interaction in T7. Tyr-404 (Rx$_3$Hx$_3$Y-DNAExG) and Tyr-409 (HLY409PSH) form the bottom of the nucleotide binding site.

The active site of B family polymerases contains a DTED motif, which, however is DTDG in the archaeal subfamily. In the pol, the carboxylate terminal Tyr-402 from the adjacent strand provides an electron withdrawl group, or a proximal coordination for metal coordination or binding of the 3′ end of the primer. The orientation of Tyr-402 is stabilized by an aromatic cluster that also includes Pro-361 and Tyr-368. Archaeal *Methanococcus vulnare* and *Thermococcus* sp. pol's have Tgo, Tyr or position 541—Phe in Tgo—rather than at 402, but might also satisfy an aromatic group. The displacement of a functional element from serine in DTDG to Tyr-402 or Tyr 541 might stabilize its conformation as an adaptation for thio-nucleotides.

The conserved cluster of amino acids DSTR, KNHG form an unexpected metal-binding site for Mn^{2+} and Zn^{2+} (Fig. 4). Its proximity to Asp-487 and to the conserved location of the dNTP γ-phosphate suggests a supporting role in nucleotide binding and/or catalysis.

3′ → 5′ Exonuclease Active Site. Tgo pol is characterized by a strong 3′ → 5′ exonuclease activity, unlike eukaryotic B type polymerases (unpublished results). The exonuclease active site is formed at the interface between the exonuclease domain and the tip of the thumb (Fig. 5). All residues important for catalysis are located in the exonuclease domain, which, at least for T4 (45), retains activity when dissociated from the polymerase (45). However, the thumb domain, with, for example, RB69 gp45 Phe-120–Arg-interaction, partially encircles the binding geometry of single strand DNA (21, 45).

The exonuclease structures of Tgo and pol3 DNA polymerases are similar in the editing site but differ considerably at the exonuclease–thumb interface. Mutant 19 restores the vestal forking DTDHE motif and notably superimposes with the equivalent strand from pol3, allowing modeling of a single-stand DNA segment into the exonuclease site based on the Dnaf/RB69-pol3-pol3/α complexes (21). The conserved residues Asp-344 and Glu-345 in the Exo I motif, Tgo 389, Asn 390, Tho 394, and Asp-355 in Exo II, and Tyr-341 and Asp-345 in Exo III are in approximately DNA-binding conformations (Fig.

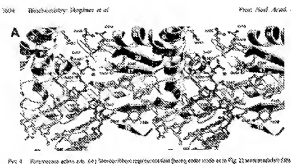


FIG. 4. Polymerase active site. (A) Stereoscopic representation (wall-eyed viewing) of the same model as in Fig. 2 with modified DNA. Amino acid residues are shown as ball-and-stick representations, with carbon (green), nitrogen (blue), oxygen (red) atoms. The DNA template (light brown), primer (light brown), and dNTP (orange) complex has been taken from the coordinates of T7 replication complex [47] for superimposing. DNA, DNA; and adjacent residues with corresponding residue in T7 are D903 and E907. Phosphorus atoms are yellow. The two metals of the T7 replication complex are shown as magenta spheres. (B) Experimentally observed metal binding site for Mn^{2+} ($K_d = 4$ mM) is coordinated as Mn and Zn^{2+} at the "low site" crystal form. The carboxylates E370 and E380 are conserved in type B polymerases (Fig. 3).

942). Moreover, the sliding shift is demonstrated by a displacement of the tip of the thumb toward the intermediate domain to grab the template strand (Fig. 5B). This shift is correlated with a large change in side chain interactions of R668 and L671 to R5N69 and Leu-1 (hydrogen in T4 gp43), the hydrogen bonds to the T base and maintains Phe-121, which intercalates between the first two bases in T4 gp43, the loop is turned backward, away from the minus, and Phe-670 (the equivalent of gp43's Tyr-127), attaches to Pro 394 10 Å away from intercalation site. This shift allows the tip of the thumb to move into the sliding channel and to block the exonuclease site.

Are There Different Conformations in Polymerase and Editing Modes? If a closure conformation of the exonuclease domain prohibits single-strand binding, an open conformation is necessary for editing. The observed distinct conformations may experiment the enzyme to "polymerase" mode. Preliminary analysis of the crystal structure of Taq pol in the low-salt conditions indicates a structural change at the interface of exonuclease and thumb, possibly reflecting a transition between open and closed forms. The altered conformation observed here may, however, be a morphometrical artifact of the high ionic strength used for crystallization. Detailed structures of the enzyme in both polymerase and editing modes are required.

Adaptation to High Temperatures. *Taq* is a thermophilic eubacteria, extremely thermophilic archaeon, with a growth range between 75°C and 98°C. For extensive replication at the temperature range, the polymerase must not only be stable, but must also adequately bind substrate DNA. A comparison with gp43 from the mesophilic bacteriophage RB69 indicates several such adaptations to high temperatures. Several loops are shortened in *Taq* pol than in gp43 (Fig. 2), and there is an increase in hydrogen bonding at several contacts. *Taq* pol secondary structure includes 40% helix, 22% B-strands, and 19% turns (calculated according to ref. 48), whereas gp43 has 42% helix, 17% B-strands, and 18% turns.

Although very strong cytoplasmic or nuclear proteins, two disulfide bridges occur in thermo- cysteine pairs 428–442 and

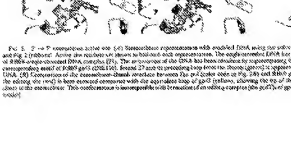


FIG. 5. T7 → P interactive active site. (A) Stereoscopic representation with exploded DNA using the color code of Fig. 4A (ball and stick) and Fig. 2 (ribbons). Amino acid residues are shown in ball-and-stick representations. The single-stranded DNA binding tunnel from the exonuclease, as B-DNA single-stranded DNA, complex [49]. The arrangement of the DNA has been simulated by superimposing the E907-N3 model in *Taq* pol with corresponding model of R5N69 gp43 (2NL166). Bound 2′ and 3′ protruding loop from the thumb (green) is apparently in collision with the modified DNA. (B) Conformation of the exonuclease–thumb interface between *Taq* pol (color code as Fig. 2A) and R5N69 gp43 (green). In *Taq* pol, the tip of the editing site (red) is here screwed connected with the equivalent loop of gp43 (yellow), allowing the tip of the thumb to move toward A (Fig. 2) close to the exonuclease. This conformation is incompatible with formation of an editing complex (like gp43), of specific shown subsites upon editing model.

106–109, although reduced, are present for recombinant (Fig. 2). Mutant cofactorless and electron density deposition with and without cofactors for the disulfide bond-reduced (by reduced state fabricated) overexpressed 561-561 domains. 2.8 Å and 3.0 Å. This is consistent with our E. coli expression and further indicates that actual perturbation by reversible oxidation. These cofactors are located in the polar domain and are conserved among B type enzymes from hyperthermophilic archaeal-methanogenic archaeons, but not among mesophilic homologs (Fig. 3). The Cys-428–Cys-442 bridge stabilizes the octapeptide fold of the loop segment between helix N in the polar domain and helix O in the flagpole and presumably aids the selective orientations of these helices. In addition, the loop segment packs against helix Q in the polar domain. Helix Q, the core of the polar domain, is the last α-helix at the first helix-turn by the second disulfide bridge between Cys-301 and Cys-305.

A much enhanced transplacentary positive potential for all three DNA-binding fields of Tpn pol is observed relative to pol1 (Fig. 2). Thus, in addition to hydrogen bonding and specific DNA-protein interactions, binding to Tpn pol has an additional strong, stabilizing electrostatic component.

An attribute in the number of salt bridges is often associated with thermostability. Although Tpn pol has a greater total number of charged residues (202) than pol1 (242), both molecules have 54 salt bridges within a 5–5 Å distance. However, at the 5–7 Å range of charge distance, Tpn pol has 77 ion pairs compared with 62 for pol1. This large increase results in a more highly charged surface of Tpn pol, accompanied by a more balanced charge distribution, compared with pol1 where charges are often located in patches (Fig. 2).

We thank Dr. Lind Jacob and Mostly Augustin for introducing Stephens, Inc., Medical Studies for help with expression of telomeric diseases vaccine, and Dr. Hsin Pan-Hou and the 560K laboratory at DESY Hamburg for help with data collection.

1. Wang, J., & Lehman, B. (1996) Annu. Rev. Biochem. 65, 721–756.
2. Mizuuchi, K., & Mizuuchi, M. (1980) Biochemistry 19, 3071–3077.
3. Wang, T. S.-F. (1991) Annu. Rev. Biochem. 60, 513–552.
4. Kaufmann, A. & Baker, T. A. (1992) EMBO Replication Symposium, New York).
5. Joyce, C. M. & Steitz, T. A. (1994) Annu. Rev. Biochem. 63, 777–822.
6. Perler, F. B., Kumar, S. & Kong, H. (1996) Adv. Protein Chem. 48, 377–435.
7. Steitzagen, C. A. & Steitz, T. A. (1989) Cold Spring Harbor Symp. Quant. Biol. 54, 143–156.
8. Arnold, E., Hong, J., Hughes, S. H. & Hostetler, Z. (1992) Chromosomal Res. 1, 19–37.
9. Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. & Steitz, T. A. (1985) Nature (London) 313, 762–766.
10. Freemont, P. S., Friedman, J. M., Beese, L. S., Sanderson, M. R. & Steitz, T. A. (1988) Proc. Natl. Acad. Sci. USA 85, 8924–8928.
11. Kiefer, J. R., Mao, C., Hansen, C. J., Basehore, S. L., Hogrefe, H. H., Braman, J. C. & Beese, L. S. (1997) Structure (London) 5, 95–108.
12. Kiefer, J. R., Mao, C., Braman, J. C. & Beese, L. S. (1998) Nature (London) 391, 304–307.
13. Kim, Y., Eom, S. H., Wang, J., Lee, D. S. Suh, S. W. & Steitz, T. A. (1995) Nature (London) 376, 612–616.
14. Doublié, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998) Nature (London) 391, 251–258.
15. Dutnall, R. N., Rhomaeus, R. J., Hosomadze, Z., Pavo, P. A. & Hunziker, R. Z. (1996) Cell 84, 1063–1068.
16. Sawaya, M. R., Pelletier, H., Kumar, A., Wilson, S. H. & Kraut, J. (1994) Science 264, 1930–1935.
17. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. & Kraut, J. (1994) Science 264, 1891–1895.
18. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. (1992) Science 256, 1783–1790.
19. Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D., Jr., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P., et al. (1993) Proc. Natl. Acad. Sci. USA 90, 6320–6324.
20. Jäger, J., Santelhan, L. J., Wang, J., Benson, G. C. & Steitz, T. A. (1994) Structure (London) 2, 869–876.
21. Wang, J., Sattar, A. K., Wang, C. C., Karam, J. D., Konigsberg, W. H. & Steitz, T. A. (1997) Cell 89, 1087–1099.
22. Ealgelf, D. B. & Eisenberg, M. (1984) Methods Enzymol. 108, 197–205.
23. Otwinowski, Z. & Minor, W. (1997) Methods Enzymol. 276, 307–326.
24. Delbaere, L. T., Vandonselaar, M. & Prasad, L. (1992) Can. J. Biochem. 70, 150–160.
25. Brünger, A. T. (1992) Nature (London) 355, 472–475.
26. Brünger, A. T. (1992) X-PLOR, A System for X-Ray Crystallography and NMR (Yale Univ. Press, New Haven, CT).
27. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991) Acta Crystallogr. A 47, 110–119.
28. Kraulis, P. J. (1991) J. Appl. Crystallogr. 24, 946–950.
29. Merritt, E. A. & Bacon, D. J. (1997) Methods Enzymol. 277, 505–524.
30. Nicholls, A., Sharp, K. A. & Honig, B. (1991) Proteins Struct. Funct. Genet. 11, 281–296.
31. Ollis, D. L., Kline, R., Steitz, T. A. & Moras, D. (1988) J. Mol. Biol. 202, 505–520.
32. Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. & Steitz, T. A. (1985) Nature (London) 313, 818–822.
33. Kiefer, J. R., Mao, C., Braman, J. C. & Beese, L. S. (1998) Nature (London) 391, 304–307.
34. Doublié, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998) Nature (London) 391, 251–258.
35. Dutnall, R. N., Rhomaeus, R. J., Hosomadze, Z., Pavo, P. A. & Hunziker, R. Z. (1996) Cell 84, 1063–1068.
36. Sawaya, M. R., Pelletier, H., Kumar, A., Wilson, S. H. & Kraut, J. (1994) Science 264, 1930–1935.
37. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. & Kraut, J. (1994) Science 264, 1891–1895.
38. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. (1992) Science 256, 1783–1790.
39. Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D., Jr., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P., et al. (1993) Proc. Natl. Acad. Sci. USA 90, 6320–6324.
40. Jäger, J., Santelhan, L. J., Wang, J., Benson, G. C. & Steitz, T. A. (1994) Structure (London) 2, 869–876.
41. Wang, J., Sattar, A. K., Wang, C. C., Karam, J. D., Konigsberg, W. H. & Steitz, T. A. (1997) Cell 89, 1087–1099.
42. Ealgelf, D. B. & Eisenberg, M. (1984) Methods Enzymol. 108, 197–205.
43. Otwinowski, Z. & Minor, W. (1997) Methods Enzymol. 276, 307–326.
44. Collaborative Computational Project, No. 4 (1994) Acta Crystallogr. D 50, 760–763.
45. Tronrud, D. (1992) TNT, Thesis (Tuskegee University, Tuskegee).
46. Adams, P. D., Pannu, N. S., Read, R. J. & Brunger, A. T. (1997) Proc. Natl. Acad. Sci. USA 94, 5018–5023.
47. Weiner, S., Kollman, P., Nguyen, D. & Case, D. (1986) J. Comput. Chem. 7, 230–252.
48. Nicholls, A., Sharp, K. A. & Honig, B. (1991) Proteins Struct. Funct. Genet. 11, 281–296.
49. Kraulis, P. J. (1991) J. Appl. Crystallogr. 24, 946–950.
50. Merritt, E. A. & Bacon, D. J. (1997) Methods Enzymol. 277, 505–524.
51. Ollis, D. L., Kline, R., Steitz, T. A. & Moras, D. (1988) J. Mol. Biol. 202, 505–520.
52. Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. & Steitz, T. A. (1985) Nature (London) 313, 818–822.
53. Kiefer, J. R., Mao, C., Braman, J. C. & Beese, L. S. (1998) Nature (London) 391, 304–307.
54. Doublié, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998) Nature (London) 391, 251–258.

Appendix 1

We have purified and characterized the Family BDNA polymerase from the archaeon Methanococcus maripaludis, cloned from ATCC 43000. This polymerase has a 41% sequence identity and 63% sequence similarity with Vent DNA Polymerase when analyzed using NCBI Blast 2 and the default parameters.

We performed the titration assay described in Example 1 of the patent application, using the Mma, Vent (exo-), and 9°N (exo+) DNA Polymerases. Experimental details and data are given in the attached figure.

For each of the three polymerases, a comparison of lanes using dideoxyCTP (ddCTP) with those using equivalent concentrations of acycloCTP (acyCTP) reveals shorter products in lanes utilizing acyCTP. These shorter products result from more efficient insertion of the acyCTP terminator compared to incorporation of the ddCTP terminator. Thus, all three polymerases incorporated acyCTP more efficiently than ddCTP.

Figure Legend

The ability of acyNTPs and ddNTPs to act as chain terminators was tested using a titration assay of the type described in Example 1. Incorporation of ddCTP was compared to that of acyCTP, respectively, using Methanococcus maripaludis DNA polymerase, 9°N (exo+) DNA polymerase and Vent® (exo-) DNA polymerases.

Incorporation of ddCTP and acyCTP was assayed by mixing 8 µl of reaction cocktail (0.025 µM 5' [FAM] end-labeled #1224-primed M13mp18, 62.5 mM NaCl, 12.5 mM Tris-HCl (pH 7.9 at 25°C), 12.5 mM $MgCl_2$, 1.25 mM

dithiothreitol, *Methanococcus maripaludis* DNA polymerase or 0.125 U/µl 9°N (exo+) DNA polymerase or 0.125 U/µl Vent® (exo-) DNA polymerase) with 2 µl of 5X nucleotide analog/nucleotide solution to yield the final ratios of analog:dNTP indicated in the figures. After incubating at 72°C for 20 minutes, the reactions were halted by the addition of 10 µl formamide. Samples were then heated at 72°C for 3 minutes and a 1 µl aliquot was loaded on a 4% polyacrylamide urea gel and detected by an ABI377 automated DNA sequencer.

ddCTP v. acyCTP incorporation by archaeal DNAPs